

P

SEARCH REQUEST FORM  
Scientific and Technical Information Center

Access DB# 54401

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
An Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone #: \_\_\_\_\_  
Searcher Location: \_\_\_\_\_  
Date Searcher Picked Up: 11/7/01  
Date Completed: 11/8/01  
Searcher Prep & Review Time: \_\_\_\_\_  
Client Prep Time: \_\_\_\_\_  
On-line Time: \_\_\_\_\_

Type of Search

NA Sequence (#) 1  
AA Sequence (#) 1  
Structure (#) \_\_\_\_\_  
Bibliographic \_\_\_\_\_  
Litigation \_\_\_\_\_  
Fulltext \_\_\_\_\_  
Patent Family \_\_\_\_\_  
Other \_\_\_\_\_

Vendors and cost where applicable

STN \_\_\_\_\_  
Dialog Reverse to NA  
Questel Orbit \_\_\_\_\_  
Or Link \_\_\_\_\_  
Lexis Nexis \_\_\_\_\_  
Sequence Systems 05  
WWW Internet \_\_\_\_\_  
Other (specify) \_\_\_\_\_

**From:** Ibrahim, Medina A.  
**Sent:** Tuesday, November 06, 2001 1:33 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 09/538, 396

please do the following search:

1. standard and interference search of a nucleotide sequence encoding SEQ ID NO:2 against all DNA databases.
2. oligo search of at least 30 contiguous bases of SEQ ID NO:1. Thanks

*Medina A. Ibrahim*  
*Patent Examiner*  
GAU:1638  
CM1-9A12  
Mailbox:9E12  
(703)306-5822

*Edward A. ...*  
*Technical ...*  
*STIC-Biotech*  
*CM 12-10-2001*

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

*The Pending database search results should not be left in the case because they contain data that is confidential.*

According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

Two new databases have been created to hold the pre-published sequences:  
**Published\_Applications\_NA** contains nucleic acid sequences; the search results will have the extension .rnpb.

**Published\_Applications\_AA** contains amino acid sequences; the search results will have the extension .rapp.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicant's submit changes to the application, they may request that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

**Sequences in the PGPub database are public information; it is permissible to leave these results in the case.**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 21:39:51 ; Search time 148.21 Seconds  
(without alignments)  
6864.174 Million cell updates/sec

Title: us-09-538-396-1

Perfect score: 4492  
Sequence: 1 aatcgagcagagtgatcc.....atctagcatcaaaaaaaaa 4492

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PTBUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	ID	Description
-----			

No matches found

Search completed: November 8, 2001, 00:54:46  
Job time: 11695 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 23:14:51 ; Search time 85.56 Seconds  
(without alignments)  
3456.429 Million cell updates/sec

Title: US-09-538-396-1

Perfect score: 4492  
Sequence: 1 aatcgcgcacgagtgatcc.....atctagcatcaaaaaaaaaa 4492

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 84510 seqs, 32917641 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending\_Patents\_NA\_New:\*

- 1: /cgn2\_6/ptodata/1/pna/PCT\_NEW\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pna/US06\_NEW\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	ID	Description
-----			

No matches found

Search completed: November 8, 2001, 02:41:12  
Job time: 12381 sec



[illegible]

1242 GTTACTAGAAACAAATAATCAGTAAACTCTGTAAAGATATGATGATGAAG 1291

334 IAsertyIserSerValLeuSerLySglnAsnSerGluLeuThHISglu 350

1292 CATCTTATAGCTCCGCTTCTGTCCAAACAAATTCGAAATTAAACAATGAA 1341

351 ILegLyLeuGlnAbISgluLAspAlaHisLeuThrMeLtySHISgl 367

1342 ATTTGGAAGACTCCAGGCAAGACTATGCTACCTCGACTTATGAAGACATGA 1391

367 uATgAspSerAspILeLyAsnILecySThrLySHIsnLeuGlyProV 384

1392 ACAGAGCTCAGACATATAAAATATATGCACTAAACATATACTTGSGCGG 1441

384 aLProGluHisProPheThrAsnAspValAlaMetAsnLeuThrAsnArg 400

1442 TTCCGTGAACATCCCTTTACGATGATGTTGCTATGACCTTACAAACAGC 1491

401 ILelYgAlAArgLeuSerSerLeuGlnsAspLeuAspLySlyLy 417

1492 ATTTAAAGCCGAGCTATCAAGCTTTGAGAAATGTTTGCTGGTATTAAGAA 1541

417 sSerAsnGlnAspGlnLeuAspValLeuThrPrySHIStryLeuLySlea 434

1542 ATTCATATGAAGATCACTTGAATGTTTGTGGGAACACTATCTTAAATATA 1591

434 snAlaArgTySerGluValAspGlyGlnILegInSerLySILeGluSer 450

1592 ATGCTGCTACTCCGAAGTTGATGGTCATATCAATCAATCAATGAATTGATCC 1641

451 MetSerGlyILeLeuArgArgArgArgLySAspLySglnArgAspAl 467

1642 ATGTCAAGGCAATTTTAAAGCGGAGAAAGATTAAGAGAAAGAACCGGATGC 1691

467 aAlaGlnValGlnLeuSerLySPhasnLeuSerArgILeAspGluArg 484

1692 TGCAGAGTGGACCTTCAAAATTTATATATCCGCTATCGATGAGAGCG 1741

484 LuArgHISMetGlnILeGluValGluArgLySThrLeuAlaLeuGlyGlu 500

1742 AGAGACATATGCAAAATTGAAGTCAGAGAGAAAGACACTTGCCTTGCAGAA 1791

501 ArgAspTyArgSerSerILeIleSerGlnLySAArgThGlnValTySerLe 517

1792 AGAGACTATATTCATTTATTAATCGAAGCAACAGAAAGTATATATGTTT 1841

517 uGlnGlnLySILeLyValLeuLeuArgGlnLySAspILeIleAsnArg 534

1842 GGAACAGAAAATTAAGTCTTCTGCGGAGAAAGTATATAATAATTATCAA 1891

534 snAlaAspGluArgValLySLeuGlyLeuLySAspAlaLeuGlnSer 550

1892 ATGCTGATGAAAGAGTAAACTGGGTGTGAAGAGGATCATTTGGAAGAC 1941

551 SerLyAspLySLeuAsnGlnILeValAsnGlnHISLySAspLySILeLy 567

1942 AGCAGAGCAACAGCTCATATGATGATGATTAATGAGCATTAAGGATTAATCAA 1991

567 sLyValLeuArgGlyArgAsnProPheGlnLySAspMeLtyLySgln 584

1992 AAAGGTCCTTAGGGGGGAGATCCTTTTGAGAGGATATGAAGAAAGGAGA 2041

584 IeaGnglnAlaPheTrpProValAspLySglnTyraSngILeuaArgSer 600

2042 TCATCAAGGCTTTTGCCCTGTGGCAAGAGAAATACAAATGATTAAGATCA 2091

601 LySerGlnGlnAlaGlnGlnGlnLeuLySPheThHISnSerLySValTh 617

2092 AAATCCCAAGAGAGAGCAAGAGCTTTAAATTTACTCAAGCAAGCAAGTAC 2141

617 rAspAlaArgGlnGlnLeuThrTyLeuArgArgAspMeLAspAlaLySA 634

2142 TGATGCTAGACAAATGACAAAACCTCGAAGATATGATGCAAAA 2191  
634 rgarphleuaspserlyseuInserIleleuGlnIleserAlaasn 650  
2192 GAAGATTCCTGGACTCGAAACTCTCAATCTATTCTACAGATATCTCTAAT 2241  
651 ValaspmethepTolysValleuGlnsPrAlaMetAsnLysArgAspG 667  
2242 GTTGACATGTTCCCAAGTTCTACAAAGCCGATGAACAAAGAGATGA 2291  
667 uGlnLysArgleuGlnuAsnPhelaAsnGlymetArgGluMetLeuAlap 684  
2292 ACAGAAAAGATTAGAAATTCGCAAAATGGAATGGGGAAATGCTTGCC 2341  
684 ropheGlnHIsleuAlaArgLysAsnHIsValCysProCysCysGluArg 700  
2342 CTTTGAACATTTGGCTCGAAGATCAATGATGCCCATGCTGTGAACCT 2391  
701 AlaPheThrProAspGluGluAspGluPheValLysLysGlnArgMetC 717  
2392 GCTTCACACCTGATGAGAGATGATGCTGTAAGAAAACAAAGATGCA 2441  
717 nasnserserThrAlaGluArgSerLysAlaLeuAlaMetGluSerSer 734  
2442 AAACCTAAGTACTGACAGAGATCTAAAGCTCGCAATGGAATCTCA 2491  
734 snAlaGlnAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleArgAsp 750  
2492 ATGCTGAACCTTTTTCAGCATGGAATGAACTTCGCACTATCTATGAT 2541  
751 AlaTyrValLysLeuValGlnLutThrIleProLeuAlaGlnLysAsnLe 767  
2542 GCTATGTAAGCTGTAGAAAACCAATCTCTAGCAGAGAAAAACT 2591  
767 uAsnGlnHIsleuAlaAspGluSerGlnLysAlaGlnAlaPheAspSer 784  
2592 GAATCAACATTTGGCGGATGAAGTCAGAAAGCGCAGGCTTTGATGATC 2641  
784 euLeuGlnValleuAlaHIsValGlnMetAspArgAspAlaValaGlnAla 800  
2642 TTTTGGGCTTCTTGCCCATGTTCAATGACAGCGATGCGATGGAAGCC 2691  
801 LeuLeuGlnProThrAspThrIleAspArgHIsValHIsGlnIleGln 817  
2692 TTATTCACACCCATGATCTATGACAGCGATGATCATGAAATTCACA 2741  
817 nleuValLysGlnValGluAspLeuGlnLutThrAlaLeuAspSerSerGly 834  
2742 GCTAGTCAAGAGATGAGATCTGAATATGCACTTGATTTAGTGCC 2791  
834 rGlnLysSerLeuGlnGlnIleGlnLeuGlnLeuAsnPheluGln 850  
2792 GAGGTCTCAAGCTTTGGAAGAAATTCAACTGAGCTGAACTTCTGCG 2841  
851 ArgThrArgAspThrLeuIleValGlnValAspAspLeuArgAspGlnH 867  
2842 AGAACACAGGACACATGATGTCGAAGTGAATGCTTAGACATCAACA 2891  
867 sarGmetLeuAsnGlnuAspMetSerSerAlaGlnValArgTrpHIsAsn 884  
2892 TAGAAATGCTAAATGAAGATGTCAGTGTCTAGGTAGATGGACACATG 2941  
884 laArgGlnLysValLysValLysSerSerIleLeuGlnuArgPheGlnLys 900  
2942 CTGGGGAAGAAAGTGAAGCTTCTGCAATTTGGAAGATTCACAAA 2991  
901 SerGlnGlnLysLeuValLeuLeuAlaGlnLysGlnLysGlnIleVal 917  
2992 TCTGAAGAGAAATGCTGCTTCTAGCTGAGGAAAACAAACATGATGT 3041  
917 lGlnLysLysLeuLeuGlnLysLysLeuAspProLeuSerLysGlnLysG 934  
3042 AGAAAGAACTTTTAGAAGAGTCTTGTGATCCATGTCACAAAGAGAAAG 3091

934 luseLeuLeuGlnLutThrAsnAlaLeuLysGlnLysLeuAspGlnL 950  
3092 AGAGCTTGTGACAGAGTAAATGCTTTGACAAAGCTGATGAAGAG 3141  
951 TyrHIsGlnLeuAlaGlnArgLysArgGluPheGlnGlnLysAspAl 967  
3142 TATCATACAGCTTCAGAAAAGAAAAGGAGGCTCCAGCAAGATCTGATGC 3191  
967 aleuGlnArgLeuAsnMetLysIleLysGlnLysLeuAspSerLysArg 984  
3192 TCTTGGAAGACTTAATATGAAGATTAAGAGGATCTGTTCCAGAAAA 3241  
984 snGlnLysLeuLysGlnLeuGlnGlnArgHIsValLeuCysHIsSerGln 1000  
3242 ACAGAAAGCTTAAGAAATGTCAGGAGGACATCTTCTTGCCATCTCAG 3291  
1001 LeuGlnSerCysMetAlaLysGlnGlnArgIleserAlaGlnLeuAsnL 1017  
3292 TTNAGAGTTGCATGSCAABAAGCAAGAAATATAGTGAATTAACAA 3341  
1017 sSerLysGlnLeuLeuGlnGlnLysGlnLysLysArgAsnIleAspAla 1034  
3342 GAGCAAGAAGTACTGACAGGCGCAGGCGCATGTAAGAAACATTTGATG 3391  
1034 sPAsnLeuLysTyrArgLysThrLysAlaAspValGlnLutThrArg 1050  
3392 ACAATCTCAAGTACAGAGAAAACAAAGCTGAATGGAACAATCTACT 3441  
1051 AspIleGlnSerLeuGlnLutArgLeuLeuSerIleGlnSerLeuSerAl 1067  
3442 GATATGATCACTTGAAGAAAGCTGCTTCAATAGATAGTGTGCTGC 3491  
1067 alIleGlnAlaAspLeuLysArgHIsSerGlnLysGlnLutLeuAsn 1084  
3492 TATAGAAAGCTGATGTAAGCCCATCTCAAGAAAAGAGAGCTTAAT 3541  
1084 erGlnLysAsnArgTrpGlnGlnLutThrLeuSerValTyrGlnSerAsn 1100  
3542 CAGAAATTTAAGAGGTCAGGCAAGAAACATCTCTTTATCAAAATATT 3591  
1101 SerLysHIsLysGlnLutLeuLysLeuSerGlnTyrLysAspIleGln 1117  
3592 TCAAGACCAACAAGAGCTTAAGCTGTCACAGTACAGATATCGGAA 3641  
1117 sarGlyThrAsnGlnPheLeuGlnLeuLysThrThrGlnMetAlaAsn 1134  
3642 GCGATATACTAAATCAATTTCTCCAGCTTAAGACAACTGAATGGCAACA 3691  
1134 yAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPhe 1150  
3692 AGGACTTGACAGATATATATCTGCTTGAACAAAGCTCTTAATGCGGTT 3741  
1151 HIsSerMetLysMetGlnLutIleAsnLysIleIleLysGlnLeuTrpG 1167  
3742 CACAGCATGAAGATGAGAGATTAATAATAATATCAAGAACTGTGCA 3791  
1167 nGlnThrTyrArgGlnAspIleAspTyrIleSerIleAsnSerAsp 1184  
3792 ACAGACATACAGAGCCAGGATATGATTAATCAATGAATTAATCTGATT 3841  
1184 erGlnGlnAlaGlnLutThrArgSerTyrSerTyrArgValValMetGlnThr 1200  
3842 CTGAGGCTGCGGACCTGATACATACAGCTACCCGCTGTGTAACAACT 3891  
1201 GlyAspAlaGlnLeuGlnMetArgGlyArgCysSerAlaGlnLysVal 1217  
3892 GGTATGCTGAGCTGGAATTCAGAGGCGCTGAGTCTGTGTAAGAGT 3941  
1217 lLeuAlaSerLeuIleIleArgLeuAlaLeuAlaGlnLutThrPheCysLeu 1234  
3942 TCTTGCTCTCTTATATACAGACTAGCACTTGGCGAAACTTTCGCTGA 3991



|||||  
746 GCTACGAGATAAAACTTCACTTAATGGAAAAATCGACAGCTT 795  
217 LlysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnGlu 234  
796 GAAAGACGACCTTATAGCTTCGTGAGAGCATTCCTCAAGATCAAGAA 845  
234 ySerAspAlaSerLysSerGlnMetGlnLeuLysGlnLysIleCys 250  
846 GGACAGAAATCTTCAAAAGTTGAGATGTGGAACCGAAACCGTGTAC 895  
251 GlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspGlnLeuArg 267  
896 AAAGTGTAGCTGAGTTCATTAATAGAAATGATGTTHAAAGATCTGAG 945  
267 GArgLeuGlnGlnIleAspIleLysAlaThrGluArgSerThrLeu 284  
946 AAAGCTCAGAGCAAGATATCAATAAACTCGGAAAGAGACACTTGT 995  
284 eUTHrGlnGlnHisGluLysLeuAlaAlaLeuSerGluGlnAsnGluAsp 300  
996 TTAAGGACACAGACAGACAGATATCGACATTCCTGAGGAAATGAAAT 1045  
301 ThrAspGluGluLeuMetGluThrGlnThrLysPheGluGluArgIleAl 317  
1046 ACTATTTGAGAAATTAAGAAATGAAAGCAAGTTTGAAGAAAGACTTGC 1095  
317 AleuLeuGluThrLysIleSerLysLeuValArgAspMetAspAspGlu 334  
1096 TCTTTTGGGAGCAAAATTCGAAAAATGAGAGAGAGATGGTTGACACAG 1145  
334 LaSerTySerSerValLeuSerLysGlnAsnSerGluLeuThrHisGlu 350  
1146 AAACATCAATTTCTCGCTTCAAAATGCAAAACCTAATATATGCTGAG 1195  
351 IleGlyLysLeuGlnAlaGlnIleAspAlaHisLeuThrMetLysHisG 367  
1196 ATCAGCAACCTACAAACAGAGCTGAGCTCATATCTCTGACAGAAATGA 1245  
367 uArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeuGlyPro 384  
1246 AAGAGATTTTACCATTCCAGACATTTTTCATTCATTAATTTAGAAATG 1295  
384 aLProGlnHisProPheThrAsnAspValAlaMetLysLeuThrAsnArg 400  
1296 TTCCTAGTACCCCTTCAGTACAGTAGTTTGAATCTCACAAATTCGA 1345  
401 IleLysAlaArgLeuSerSerLeuGlnAsnAspLeuLeuAspLysLys 417  
1346 ATAAATCAAGACTGGGTAACCTTGAGATGATTGCTGATTAAGAGAA 1395  
417 sSerAsnGluAspGlnLeuAspValLeuThrPheLysHisTyLeuLysIle 434  
1396 ATCAAAATGAAAGCTCTTAAGTACGGCTTGCGATTCATATAGAGCTTA 1445  
434 snAlaArgTySerGluValAspGlyGlnIleGlnSerLysIleGluSer 450  
1446 ATGACCGCTTGAAAGACATCGAAGCTCGAAGAACGACTTAAGATGAATC 1495  
451 MetSerGlyIleLeuArgArgArgLysAspLysGlnLysGluArgAspAl 467  
1496 AAGATGGCATATCAAAACGATAGAAAGAAAGAAATTAAGCGCTGATTC 1545  
467 aAlaGluValGluLeuSerLysPheAsnLeuSerArgIleAspGluArg 484  
1546 ATTTGAATTCGAAATTCCTACCTGTGATGTTAAACAACATGATGAAGAG 1595  
484 LuArgHisMetGlnIleGluValGluArgLysThrLeuAlaLeuGlyGlu 500  
1596 AGAAACAGAGTGAAGTTCGAGTTCGAGAAAGACAAACAGAAATTCGAA 1645  
501 ArgAspTyTrpSerIleIleSerGlnLysArgThrGluValTySerIle 517

1646 AGGGATTGAAATCCAAATAGAACAGAAACGATGAAATATACAGCTT 1695  
517 uGluGlnLysIleLysValLeuLeuArgGluLysAspIleIleAsnArg 534  
1696 GGAGCATTAATCAAGACACATCGAATCGAAGAAAGACCTTATGCTGTG 1745  
534 snAlaAspGluArgValLysLeuGlyLeuLysAspAlaLeuGlnSer 550  
1746 ATGCTGAGATAGAGTGAATATGCTTTAAAGAAACGAAACAGAGATAT 1795  
551 SerLysAspLysLeuAsnGluIleValAsnGlnHisLysAspLysIle 567  
1796 CTTAAGAAAGAACACAAAAGTAAATAGATGATGCAAGGATCGAATAG 1845  
567 sLysValLeuArgGlyArgAsnProPheGlnLysAspMetLysGluI 584  
1846 AGGAGTGGCGAAGGAGATTTCCACCCGAGAAAGATATGAAGAGGAAA 1895  
584 LeAsnGlnAlaPheThrProValAspLysGluTyTrpAsnGluLeuArgSer 600  
1896 TAGTTCACACTTTGAGGTCAATTTGAAGGAAATATGATGACCTCAGTTTG 1945  
601 LysSerGlnGluAlaGlnGlnLeuLysPheThrGlnSerLysValThr 617  
1946 AATCCCGAAGACAGAAAGAGGTCAACATGTCGACATGAGATGACA 1995  
617 TrpAlaAspGluGlnLeuThrLysLeuArgArgAspMetAspAlaLeu 634  
1996 AGAAATCAACAAATAGCTTTCACAGCATTAACAGATTAACAGATCAAGA 2045  
634 rArgPheLeuAspSerLysLeuGlnSerIleLeuGlnIleSerAlaAsn 650  
2046 AGAGATACATTTGATGCTCAAGCTTCAGCGCTTAACAGAAATCAGTTACC 2095  
651 ValAspMetPheProLysValLeuGlnAspAlaMetAsnLysArgAsp 667  
2096 ATGATGCTTATCCCAATTTGTGGAGTGGCTTAAGAGAAAGAAATGA 2145  
667 uGlnLysArgLeuGlnAsnPheLysAsnGluMetArgGluMetLeuAla 684  
2146 CCGTAAACGCAATATATACCTGGCAATGGATATGAGCTCAAGATTTGAA 2195  
684 rPheGlnHisLeuAlaArgLysAsnHisValCysArgProCysGluArg 700  
2196 CTTTGAAGAAACGACACATCGAAGACATTCCTGCTGCTGCTGAGAGCC 2245  
701 AlaPheThrProAspGluGluAspGluPheValLysLysGlnArgMet 717  
2246 TCTTTAAGCTGATGAGAGGCTAGCTTATTAAGAAAGGAGGTAA 2295  
717 naAsnSerSerThrAlaGlnArgSerLysAlaLeuAlaMetGluSerSer 734  
2296 GGCTTCAAGCAGAGGAGCATCTTAAGCATTAAGCATGAGATCTTCCA 2345  
734 snAlaGluAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyAsp 750  
2345 ATGCTGACTGTGATTTCAACACATGACAAACATAGCGGAGTTTGA 2395  
751 AlaTyValLysLeuValGluGluThrIleProLeuAlaGluLysAsnIle 767  
2396 GAATACCTGAAACATACTAGAAATATATCTCTGCTGAGAAACTTT 2445  
767 uAsnGlnHisLeuAlaAspLysSerGlnLysValGlnAlaPheAspArg 784  
2446 GCAAGACATTAACAGAAAGCTGGCTGAAAGTGAAGCTTGTATGATG 2495  
784 eUleuGlyValLeuAlaHisValGlnMetAspArgAspAlaValGluAla 800  
2496 TATTAGCAATCTCGCAGAGATAAAGCTGCAAGAGCTGATGAAGCA 2545  
801 LeuLeuGlnProThrAspThrIleAspArgHisValHisGluIleGln 817  
2546 CTGCTGACGCCCTCGAAATATGCTGACAGATTTCTTACGAAATCGTATC 2595

817 nleuValLysgluValgluAspLeuGluTyrAlaLeuAspSerSerGlyA 834  
 2596 CTACCAAAAGCAAAATGAGATCTTGAAATTAACCTTGATTTAGGGGTC 2645  
 834 rglValLysSerLeuGluGluLeuGluGluLeuAsnPheLeuGlu 850  
 2646 TTGGGCTTAAACATATGAAAGAAATCCATGAGATTAAGTACGCTCCA 2695  
 851 ArgThrArgAspThrLeuLeuValGluValAspAspLeuArgAspGlnH 867  
 2696 AGCAGCAAGATTAATATGACGCTGAGCAAACTTGAAGATGACCA 2745  
 867 sArgMetLeuAsnGluAspMetSerSerAlaGlnValArgPheHisAsn 884  
 2746 GATTACCTGGAACGATATATCATGCTCCAGGCCAGATGATGATGAC 2795  
 884 LaArgGluGluValLysAlaSerSerIleLeuGluArgPheGlnLys 900  
 2796 TAAGGAGAGAAAGCCAGGAGCCCAATTTGACGGAGCTTACAAG 2845  
 901 SerGluGluGluLeuValLeuLeuAlaGluGluLysGluGluLeuLeu 917  
 2846 GCAGAAAGAGACCTGAGCGCTTACCGAGGAGAAAGTCAACTTGACT 2895  
 917 GluLysLysLeuLeuGluGluSerLeuAspProLeuSerLysGluLysG 934  
 2896 CGATGTGAAGATTTGACCGAGGCCCTGGGCCCTTATCCAAAGGAAAG 2945  
 934 LysSerLeuGluGluGluTyrAsnAlaLeuLysGluLysLeuAspGluG 950  
 2946 AGCATTTCTGATGATCATCAATGATATGAAGATTAAGCATCAACAG 2995  
 951 TyrHisGluLeuAlaGluArgLysArgGluPheGluGluGluLeuAsp 967  
 2996 TATGAAGAACTGGCTGAGAAAGAAAGAACTCAACAGAGAACTGAGGC 3045  
 967 AlenGluArgLeuAsnMetLysIleLysGluTyrLeuAspSerLysLys 984  
 3046 ATTACTCAAGGCCAGTATTAATTAATGATATCATGATCTGAAGAAAG 3095  
 984 snGluLysLeuLysGluLeuGluGluArgHisValLeuCysHisSerGln 1000  
 3096 GGGAGCGCTTGAATGATATTCAGAAAGCAAGCGGTTATCGATTCAG 3145  
 1001 LeuGlnSerCysMetAlaLysGluGlnArgIleSerAlaGluLeuAsn 1017  
 3146 CTTCAGACCTGTGAAGCTAAGAAAGATGAACTGACGCTGAACAG 3195  
 1017 sSerLysGluLeuLeuGluGluGluGluGluLysArgAsnIleAsp 1034  
 3196 AAATTAACACTTGATCGGAATCAAGATCACTGAAGAAATATGAGG 3245  
 1034 sPsnLeuLysTyrArgLysThrLysAlaAspValGluGluLeuThrArg 1050  
 3246 ATTAATTTAATCTACCGCAACAAAGCAAAAGTGAAGAGACTTACAG 3295  
 1051 AspIleGluSerLeuGluGluArgLeuLeuSerIleGlySerLeuSerAl 1067  
 3296 GAGATTGATCATCTGGAAGCAAAATATTTGATGTTGGATGGCCGC 3345  
 1067 AlleGluAlaAspLeuLysArgHisSerGluGluLysGluArgLeuAsn 1084  
 3346 AGTTTAACCTGAAATTTGAAGATATTGCGGGAAGGAAAGACTCTTT 3395  
 1084 ergLysPheAsnArgTrpGlnGluThrLeuSerValTyrGlnSerAsnIle 1100  
 3396 CAGACTGACCGGTGCTGGAACAGTCTCTTTACAGAGCAAGCTAT 3445  
 1101 SerLysHisLysGluGluLysLeuSerGluTyrLysAspIleGluL 1117  
 3446 TCAAGAACAGAGTGGAGCTAAACAGCACAATTCAGAGCAATGATTA 3495

1117 sArgTyrThrAsnGlnPheLeuGluLeuLysThrThrGluMetAlaAsn 1134  
 3496 GCGGCACTTTGATCACTGATCCAGTAAAGACAACTGAATGACCAATA 3545  
 1134 yAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPhe 1150  
 3546 AAGACTTGATGATCACTACAAATGCCCTTACAAAGCAACAAATGCCCTC 3595  
 1151 HisSerMetLysMetGluGluLeuAsnLysIleIleLysGluLeuTyrG 1167  
 3596 CACCAATGAAATGAGAAATTAATTAAGATTATTAAGGAGACTGTGCA 3645  
 1167 ngLThrTyrArgGluGluAspIleAspTyrIleSerIleAsnSerAsp 1184  
 3646 GCAGACATACAGAGCTCAGATATGATTCATCATCAATACACACAGATT 3695  
 1184 ergLysAlaGluLysThrArgSerTyrSerTyrArgValValMetGlnThr 1200  
 3696 CCGAAGGTGACGCCGCCGCTTACAGCTACAGAGTTCATTATCCAGACT 3745  
 1201 GlyAspAlaGluLeuGluMetArgGlyArgCysSerLeuGluGluLys 1217  
 3746 GGTATACAGAACTTGAATGAGAGAGATGACAGTCAGGTCAAAAGT 3795  
 1217 LLeuAlaSerLeuIleLeuArgLeuAlaLeuAlaGluThrPheCysLeu 1234  
 3796 TCTTGCTTCATGATATTAATGATGCTTACCGAGACATTTGCTTA 3845  
 1234 snCysGlyIleLeuAlaLeuAspGluProThrThrAsnLeuAspGlyPro 1250  
 3846 ACTGGAAATTTAGCCCTTGATGAGCAACTACAAACCTGGATGGGCC 3895  
 1251 AsnAlaGluSerLeuAlaAlaAlaLeuLeuArgIleMetGluAlaArg 1267  
 3896 AATTCAGAAATCTTTCGAGAGCTCTTTCATGATCATGAGAGACAGAA 3945  
 1267 sGluGluGluAsnPheGlnLeuIleValIleThrHisAspGluArgPhe 1284  
 3946 GGGCAGAGAAATTTCCACATCATGATCATCCAGATGAAGCTTTG 3995  
 1284 LHisLeuIleGlyGlnArgGluLeuAlaGluLysTyrTyrArgValSer 1300  
 3996 CTCAAATGATGGCCCAAGCAACATGCTGAAAATATTAACCGGCTGCA 4045  
 1301 LysAspGluAsnGlnHisSerIleIleGluSerGluGluIlePheAsp 1316  
 4046 AAAGATGATATGCAACAGCATATGAGGCCCAAGAGATCTTGAT 4093  
 seq\_name: gb\_p11:AC006223  
 seq\_documentation\_block:  
 LOCUS AC006223 100567 bp DNA PLN 05-APR-2000  
 DEFINITION Arabidopsis thaliana chromosome II section 180 of 255 of the  
 complete sequence. Sequence from clones F20M17, F22D22.  
 ACCESSION AC006223 AE002093  
 VERSION AC006223.3 GI:6598534  
 KEYWORDS  
 HIG.  
 SOURCE  
 ORGANISM  
 thale cress.  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE  
 1 (bases 1 to 100567)  
 Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,  
 Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,  
 Feldguy,M.T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,  
 Koo,H., Moffat,K.S., Cronin,L.B., Shen,M., Vanaken,S.E., Umayam,L.,  
 Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Cressy,T.H.,  
 Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Pruss,D.,  
 Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and  
 Venter,J.C.  
 TITLE  
 Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana





## alignment\_scores:

Quality: 3050.00 Length: 2727  
 Ratio: 2.661 Gaps: 45  
 Percent Similarity: 42.024 Percent Identity: 30.436

## alignment\_block:

us-09-538-396-2 x AC006223 ..

Align seg 1/1 to: AC006223 from: 1 to: 100567

```

1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPheAspPr 17
17700 ATGAGTACGGTCGATTAATGTTGATCAAGGATATAAGAGCTTCGACCC 17749
17 0AspAsnLysAsnValIleThrPhePheLysProLeuThrLeuIleValG 34
17750 AGAAATATAAGAAAGCTGTTACTTCTTCAGACCTTAAACCTTAACGTCG 17799
34 1YProAsnGlyAlaGlyLysThr..... 41
17800 GCGCTACGAGACCGGAAAGAACTGTAAGTCTTAAACCTTTCTCTTC 17849
41 ..... 41
17850 TCCCTCAGTTCACTCCCATTTCTTCGCCCTCCGTAATTGAAGAGTTGTGAC 17899
42 ..... 52
17900 TATTTCTGATTTTTCAGACGATTAATAGAGTGTAAAGGTTGTTGTA 17949
52 hrgIgluLeuProPheAsnSerArgSerGlyHisThrPheValHisAsp 68
17950 CCGGAGAGTGGCCACGAAACGAGAGTCTGTCATACGCTTCATTCATGAC 17999
69 ProLysVal..Ala..... 72
18000 CCTAAGGTAGTCTCTAGTTAGATTCCTCAATTCCTAATTTCAATGTT 18049
72 ..... 72

```

```

18911..19025,19249..19477,19659..19837,19925..20095,
20187..20291,20381..20488,20388..20740,20875..20978,
21144..21271,21551..21665,21785..22033,22232,
22429..22637,22749..22950,23535..23704,23764..24100,
24481..24606,24692..24744,24855..25005,25081..25143,
25254..25387,25590..25722)
/gene="At2g31970"
/codon_start=1
/product="putative RAD50 DNA repair protein"
/protein_id="RAD15407.1"
/db_xref="gi:4263721"
/translation="MSTVDKMLIKGIRSPENKNVYFRRLTLIVGANGAKTII
ECLASCTGELPNRNSGSHFTHDPKVAETEKQKIKRFTAGKADVVCIRFDT
QKASKMEYKAISSVLTINPHTGEKVCISYRCADNRELIPALMGSKALLENITFHO
DESNMPLDPSLKKKFDIFSATRYTKALEYIKKLHKQAOEIKTKFLKLELTQTK
DAAYKLREIAQOERTESSKQOMLETVOVDVAENKEMIKLMDLKLQDOVISIK
TAEKSTLFEKQOYALPEENEDTIEELKEMKSEERLALIGTKIRMERMVTE
TTSSTAHNAKTNMLEISKLOTEAAHMLKNRSDTIONIPHYMLGVSPPESTE
VVALNTRIKSRIGLEMDLDKSKSNETALSTANDCYDANDNRKSTINOKRANEI
KMGISRIEKEIERDSFEFEISTVDVKOTDEREKQVVELEKTKQNSERGEKTE
OKQHEYSLEHKIKTLNREDDYAGADNRLRLIDECDRLRGVLKGLPPEKDKR
EIVQALRSIEREDDLSLKREAEVNNLMKIQVNNLSFKHNDTESRKRYIESK
IQAIOESYTDVAPKLESADKDRDKREYNNAMGOMPEPFKRAKROESCC
ERSTDEASPFKROKRVKASSTGEHLKALAYSSNADSVFOLDKRLAVFEESLT
TEIPLAEKTLQEHTELOKSEALDVGISQIQRADSDSTALVQPLENDRTOE
IYSYQOIEDLEKIDFRLGVKTEIEIOSESSLOSCKDLGELKELKDDQITER
DISCLOARHAAVEERAKAANLVDYTKAEEDEKRYNSQDLVDVKTLYTEALGLS
KEKEQLSDYNDKIRRNQVEIELAKRKNRYOQEVALLKASKINDCTRYDLKKG
ERLDDIOEKORISDQLOSGEARKNELAGELNNKRNODLRINIDNLNRPYTK
AKVEELTRTESLEEDILINIGCAIAYEALIVKILREBELSLNCRGTVSYESSI
SKRVELKQKHQKIDKRNHDQLOLQKTEMANKDDRYNALDKRLMFRHYKMEI
NKIRLMQOYTRGDMDYIRIHSDEGATRSYKVLMTQSDTELEMGRCSAQOK

```

```

18050 TCTGTGTGATGATGATGGAATTCAGGATTCGATTTGCCCTATAGTAG 18099
73 .....GlyGut 75
18100 TATTCAGATGAAATTCATTGTTCTCCACTATATAGGTTGCTGAGAGAA 18149
75 hrgIuThrLysGlyGlnIleLysLeuArgPheLysThrAlaAlaGlyLys 91
18150 CCGAGACAAAGACAAATTAAGCTGAGATTCAGACAGCTCCAGGAGAG 18199
92 AspValValCysIleArgSerPheGlnLeuThrGlnLysAlaSerLys 108
18200 GATGTGATATATATAGGATTCATTGAGTTCAGCGAAAGGCTCAAAAGT 18249
108 tgluPheLysAlaIleGlnSerValLeuGlnThrIleAsnProHisThr 125
18250 GGAGTACAAAGCTATATGAGATGCTTCAAAACATTAATCTCAGACTG 18299
125 1Ygu..... 126
18300 GAGAGGTTGCACCTTATCTCTGTTATTATCTTCATTGGTTCTTCTTC 18349
126 ..... 126
18350 TCCCTTTTTCAGTATGATATAGATCTGATTTTGGTTGAAGGTCTC 18399
126 ..... 126
18400 GAGCTAAGTGTAGACTATCTATATTCCTACAAAGTTCAACGTTTCTT 18449
126 ..... 126
18450 TGTCTCATATAGAAATGATTGATACCTGTTTTTGAAGAAAGCTTGA 18499
126 ..... 126
18500 ACGCAGCCTCATTTGAGATTCTTAATGTTGTTGTTATTTATTCCTCC 18549
126 ..... 126
18550 ATTTTTCATGAGTCCATCATGATCATTTCTTCTTATTTTGTCC 18599
127 .....Lys..ValCysLeuSerTyrArgCysAlaAs 136
18600 ACATACATTGTCMAATTCGAGAAAGTATGCTCAGCTCAGAGTGGCTGCA 18649
136 pmetAspArgGlnIleProAlaLeuMetGlyValSerLysAlaValLeuG 153
18650 CATGATAGGAGATACACAGCTTGTATGGGTGTCCAAAGCTATTTTNG 18699
153 1uAsnValIlePheValHisGlnAspGlnSerAsnTrpProLeuGlnAsp 169
18700 AAACGATATATTTGTGCATGAGTGAATGCTAACGCGCCACTGCAAGAT 18749
170 ProSerThrLeuLysLysPheAspAspIlePheSerAlaThr..... 184
18750 CCTTCTACTTTGAAAAAGAGTTGATGATATTTCTCAGCCACCAAGTA 18799
184 ..... 184
18800 GTTTCTGGCTACCAACAATTTCTTAAATCATCATGTTCTGCTGCT 18849
184 ..... 184
18850 ATTTGTTAGTAAAGAAACCGAATCCAGAAATGTAAATGTGTGTTTA 18899
185 .....ArgTyrThrLysAlaLeuGlnValIleLysLysLeuHisLys 198
18900 ACATGAAACAGATATACAAAGGCTGTGAGATTAATTAAGAACTTCACCA 18949
198 sAspGlnMetGlnGlnIleLysThrPheArgLeuLysLeuGlnAsnLeuG 215
18950 GGATCAGCTCAGAGATTAATAAATTTCAAGTTAAATTTGGAATATGTC 18999

```

360	.....	Alah1sleuthrMetLysH1sGLua	368
19900	CATGTTTCTTATATATTTTCATAGCGTCCTCAATGCGTGAAGAAATGAA	19949	
368	rgAspSerAspIleLysAsnIleCysThrLysH1sAsnLeuIyProVal	384	
19950	GAGATTCTACCATCCAGAACATTTTTCCTATACATAATTTAGAAATGTT	19959	
385	ProGluH1sProPheThrAsnAspValAlaMetAsnLeuthrAsnArgTl	401	
20000	CCTACTACCCCGCTTCAGTACTGAGTAGTTTGGAAATCTCACAAATCGAAT	20049	
401	eLysAlaArgLeuSerSerLeuGluAsnAspLeuAspLys..Lys...	416	
20050	AAAAACAAGCTGGGTGAACCTTGATGGATGGTTGCTCGATAAGAAGGTTCC	20099	
416	.....	416	
20100	ACATTCTTATTTTCTACAGAAATTTAATGTATATTCATTATTTGGCAT	20149	
417	.....	LysSerAsnGLua 421	
20150	ATGTTGACATCTTTAGAGATTGTGATGTGCATCTCAGAAATCAAAATGAAA	20199	
421	spGluLeuAspValLeuthrPlySH1sTYrLeuLysIleAsnAlaArgTYr	437	
20200	CTGCTCTAAGTACGGCTGGGATTCCTATATATGATGTATATACCGCTGGG	20249	
438	SerGluValAspGlyGluIleGlnSerLysIleGlu..Ser.....	450	
20250	AAAAACATCGAAGCTCGAGAAACGACTAAAGATGAATCAATCAAGCTGTGTT	20299	
450	.....	450	
20300	CACCATTTGGAACTCTAACCCAAATCTCCACTAAAGAGCTTAAGTTTC	20349	
451	.....	MetSerGlyIleLeuArgAla 458	
20350	ACTCTTTTACTTTTATGATGGTGACTTATGCGAGATGGGCAATCAAAACGGA	20399	
458	rgLysAspArgGluLysGluArgAspAlaAlaGluValGluLeuSerLys	474	
20400	TAGAAGAGAAACAAATTTGAGCTGATCTCTTTGAAATTCGAAATTTCCACT	20449	
475	PheAsnLeuSerArgIleAspGluArgGluArgH1sMet.....	487	
20450	GGGATGTTAAACAACAACATGATGAAGAAGAAACAAGTGTATGCTTC	20499	
487	.....	487	
20500	GTCCTTTATTTACTTACATGATTTTGGCTTATGATGACTTTTCCCTGGCT	20549	
488	.....	GlnIleGluVal 491	
20550	AATATGCTTACTTTTTCATCATCTGCGGTTGACACAGCAAGTGGCTT	20599	
492	GluArgLysThrLeuAlaLeuGlyGluArgAspTYrAspSerIleIleSe	508	
20600	GAGAGAAACAAACAACAGATCTGAAAGGGATTTGATTCACAAATATA	20649	
508	rgLysArgThrGluValTYrSerLeuGluGlnLysIleLysValLeuL	525	
20650	ACAGAAACAGCATGTAATATACGCTTGAGCATTAATAACAAGACACTGA	20699	
525	euhArgGluLysAspIleIleAsnArgAsnAlaAspGluArgValLysLeu	541	
20700	ATCGAGAAAGAACGTTAAGGCTGCTGATGCTGAGGATGACGTGAATATG	20749	
542	GlyLeuLysLysAspAlaLeuGlnLysSerSerLysAspLys..LeuAsnGlu	557	
20750	TCTTTAAAGAAACGGAACAGAGAAATCTTAAGAAGAAACAACAAAAGAT	20799	
558	ile.....	558	

```

20800 GTGATCTCACTCACCACAACTTTCTCTGCTTTCTCTTCTTCTAT 20849
559 .....ValangulhisL 563
20850 TATCTATGTTTCTTCTCTCTAAAGCTACTTACGAAATAGATGTCGA 20899
563 ysaplysllelyslvalleuargllyarganprophglnlyasp 579
20900 AGGATCGAATTAGAGAGCTGCTGAAGGGAGATTACACCCGAGAGAGAT 20949
580 Metlyslsaglileasnglnala..... 587
20950 ATGAAGAGGAAATAGTTCAGCTTGTGAGTCGCTTTCTCACTTAAGTA 20999
587 ..... 587
21000 TTATTTTGATGATTCGAAATGAATCACTATTAATATAGATCTCTGA 21049
587 ..... 587
21050 TCAGCATCTCAAAACAGAGATAAGTCTGCTACTGTTGTTCAAGTTAT 21099
588 .....PheTProva 591
21100 CACATCTCAAAACATTCTTCTTATTGTTTCTCTCTTGGTTAGTCAAT 21149
591 lAsplysgluTyraengluleuargserlyserglnlualaglucng 608
21150 TCAAAAGGAATATGATGACCTCAGTTGAAATCCGAGACAGACAAAGG 21199
608 luleuLysPheThrglnserlyslvalThrAspAlaArggluLnuThr 624
21200 AGGTCAACATGTTGCAGATGAAGATACAGAAAGTAAACAATAGCTATTC 21249
625 LysleuAargAspMetasp..... 631
21250 AAGCATACACAGGATACAGAAATGTAGACACGTGTTAATTAAAGTTA 21299
631 ..... 631
21300 GATATCTATGCTGAGTCTCTGACTAGCGCCTTAAGTTGTGAATAACT 21349
631 ..... 631
21350 GCTCTTGACAAATTTTTCAGTGTACAAATATTAGAACATGTTAATA 21399
631 ..... 631
21400 ATGTCTCATCTGTACCCCTTTTATTGATTTGCTTCTACAAATGTACC 21449
631 ..... 631
21450 TGGATTCACATTTAAAGACAGATAGGAATCATTTTGGGTGTACAGTG 21499
631 ..... 631
21500 ATGATTAGAAAATATCTATTCACATCTGCAATCTGACTTGAATTTTATGA 21549
632 AlAluysAargArgPheleuAspserlyslenglnserlleuLglnIle 648
21550 GCAGAGAAAGATACATTTGATGTCAGCTTCAGCCTTAAACACAGAAATC 21599
648 rAlAsnValAspMetPheProLyValleuLglnAspAlaMetAsnLysA 665
21600 AGTTACCATGATGCTTATCCCAAAATGTTGAGTCGGCTAAGACAAATA 21649
665 rAspAspLglnLysArgLeu..... 671
21650 GAGATGACCGTAAACGCTTGTGTTAAACGTGGGCACTGTTTCATATCC 21699
671 ..... 671

```

---

```

21700 CCACCCCTGTATGTTGGTGAAATCCAAATATATCTCTCTTGCACTG 21749
672 .....Glu..AspPheAl 675
21750 AAAAGCATAGTCTGTTCTGCTGCTGCTGTTGACCGCAATATTAACATG 21799
675 aAsnglyMetArggluMetleuAlaProPhegluHisLeuAlaArgLysA 692
21800 AAATGCTATGGTCAGATGTTGAACCTTTGAGAAACAGACAGCTCAAG 21849
692 snHisValCysProCysCysgluArgAlaPheThrProAspGluLysP 708
21850 AGCAATCTTGTCTGCTGCTGCGAGCGCTCTTTACAGCTGATGAAGAGCT 21899
709 GluPheValLysLys..... 713
21900 AGCTTATTAAGAAGGTATACAAATCTTCTCTGTTTGGCTTAACGAT 21949
713 ..... 713
21950 TCAAAATTTTCTGCTTCTGAAATTAACGGAATGCTCTTATTGAGTA 21999
714 .....GlnArgMetGlnAsnSerSerThra 722
22000 CGTAATTTCTTTCAATACAGCAAGGTAAGGCTTCAAGCACAGG 22049
723 GluArgSerLysAlaLeuAlaMetGluSerSerAsnAlaGluAlaLeuPh 739
22050 GAGCATCTTAAAGCATTTGACAGTGGAGCTTCCAAATGCTGACTGTTAT 22099
739 eGlnGlnLeuAspLysleuArgThrIleTyraSpAlaTyraValLysLeuV 756
22100 TCACACGCTAGACAAACTACAGCGCGAGTTTGAAGAAATACTCGAAACTRA 22149
756 aLgluGluThrIleProLeuAlaGluLysAsnLeuAsnGlnHisLeuAla 772
22150 CTACTGAATATTATCTCTGCTGAGAAACTTTGCAAGACATACAGAA 22199
773 AspGluSerGlnLysAlaGlnAlaPheAspAspLeu..... 784
22200 GAGCTGGCTCAGAACTGGAAGCTCTGTATGATGTGAGCTGGGATCACT 22249
784 ..... 784
22250 TGTATTTCTTGGATAGAACTGATCCAGCTGCAATTCATATCAGTTT 22299
784 ..... 784
22300 CTTTACAAGTGCAAGGTGTGCTGTTTCTGCTGCTGTTCTCTTA 22349
784 ..... 784
22350 ATCCCTGTTGTCACATGACCTGGAATTTTAAATTCACAGTTCATAGCA 22399
785 .....Leu.GlyValLeuAlaHis 790
22400 TCTCAATCATATATGATGATGCTTTCAGGATTTAGGATCTCTCACAG 22449
791 ValGlnMetAspArgAspAlaValAlaLeuLeuGlnProThrAspPh 807
22450 ATAAAGCTGCAAAAGACTCGATGAGACACTGTGACGCCCTCGAATA 22499
807 rLLeAspArgHisValHisGluIleGlnGlnLeuValLysGluValGluA 824
22500 TGCTACACAGGATCTTTCAGGAATGTTATCTTACCAAAAGCAAAATGAG 22549
824 sPLeuGluTyraLeuAspSerSerGlyArgGlyValLysSerLeuL 840
22550 ATCTTGAATATATAGCTTATTTAGGGGCTTGGGGTTAAACATATGGA 22599
841 GluIleGlnLeuGluLeuAsnPheLeuGlnArgThr..... 852
22600 GAAATCAATCATGAATTAAGTACCTCCAAAGCAGCAAGTAGTACTGTG 22649

```



```

24450 TGCCTGAATGATTAATTTCTTCTGACGACCGGTGCGTGGA 24499
1092 hrlleuSerValTYRGLInSerAsnIleSerLysHsLysGlnIleuLys 1108
24500 CAGGTCTGCTTACGACGACGATTTCAAGAAAGAGAGGCTAAAA 24549
1109 leuSerGlnTYRlysAspIleGluLysArgTYRThrAsnGlnPheLeuG1 1125
24550 CAGGCAAAATACAGACATAGATAGACGGCAGCTTGATCACTGATCA 24599
1125 nleu.Lys 1127
24600 GCTAAAGTAAGACGTAGATATACCTTGCTGAGTACCTGAATGCTAAC 24649
1128 .....ThrThrG1 1130
24650 AAATCCACTTATGTAATTAATGACATTTCAATTTGGCAGACAACTGA 24699
1130 uMetAlaAsnLysAspLeuAspArgTYRThrAlaLeuAspLys..Al 1146
24700 AATGGCAAAATAAAGACTGGATAGATCTACAAATGCCCTTGACAACTAG 24749
1146 B..... 1146
24750 CTCTCGAGTCTTATATATACCTTCTGCGTACGTTGTTTCATTTTCAC 24799
1146 ..... 1146
24800 TATCTAATTAACAAATCATTTATCTGACGTTTAAACATATATTCCA 24849
1147 .....LeuMetArgPheHisSerMetLysMetGluIleAsnLys 1160
24850 TACAGACACACTAATCGCTTCACACATGAAATGAGAGAAATATAA 24899
1160 sIleIleLysGluLeuTPGInIleThrArgGlyGlnAspIleAspT 1177
24900 GATTATTAAGGAGACTGTGGCAGCAGACATACAGAGTCAGAGATATGAT 24949
1177 yrlleSerIleAsnSerAspSerGlnLysAlaGlyThrArgSerTyrSer 1193
24950 ACATACGAAATACACTCAGATTCGCAAGGTGACACCGCTCTTACAGC 24999
1194 TyrArg..... 1195
25000 TACAGGATATCTACAGAAATGACACATACTTACGATGTGATCATCAT 25049
1196 .....ValValMetGlnThrGlyA 1202
25050 TACGATCGTGTCAACCTGTTTCTTCTAGGTTCTTATGCGACACTGGTG 25099
1202 sPALaGluLeuGluMetArgGlyArgCysSerAlaGlyGlnLysVal..L 1218
25100 ATACGAACTTGAATAGAGAGAGATGACACTGCGACGTCAAAAGTGAAT 25149
1218 eu..... 1218
25150 TGTCTAAATCGTCTCTGTAAGACATAGATTACAGTTCATATATTATT 25199
1218 ..... 1218
25200 TAAAGCTGCTGTGCACTCAGGTCATCACTTCTTCTTCATCATCTCCC 25249
1219 .....AlaSerLeuIleIleArgLeuAlaLeuAlaGluThrPheC 1232
25250 GAGGCTGCTGCTTCAATGATTAATAGTTGGCTTTACCCAGACATATT 25299
1232 yLeuAsnArgGlyIleLeuAlaLeuAspGluProThrThrAsnLeuAsp 1248
25300 GCTTAACCTGTGGAATATAGCCCTGTGATGAGCCAACTAACCTGAT 25349
1249 GlyProAsnAlaGluSerLeuAlaAlaLeu..... 1259

```

```

25350 GGGCCCAATTGACAAATCTCTTCGAGAGAGCTCTTCTTAGTAAGACTTCAG 25399
1259 ..... 1259
25400 ATCTATACCCCTAATGATGATGATGATGATGATGATGATGATGATGAT 25449
1259 ..... 1259
25450 TTTCCTCTTACATAGACATTTAACTCTCTTAACACCTGGAGTCTATTATGT 25499
1259 ..... 1259
25500 GACTTCCTGCTCTTCCGTAATATGTTGTATACATCTTAAGACCATG 25549
1260 .....Leu.ArgIleMetGlu 1264
25550 CTCTCAACATTTGATCATATATATGCTGTTATCTTCACAGATCATGGA 25599
1265 AlaArgLysGlyGlnGluAsnPheGlnLeuIleValIleThrHisAspG1 1281
25600 GACGAAAGGGCCAGAGAAATTTCCACTCATGCTCTTCCACGATGA 25649
1281 uArgPheAlaHisLeuIleGlyLysArgGlnLeuAlaGluLysTyrTyrA 1298
25650 ACGTTTCTCTCAATGATGTCGCCAAAGCAACATGCTGAAATAATATACC 25699
1298 rGValSerLysAspGlu..... 1303
25700 GGGTCGCAAAAGATGATATGTAGCAAACTATATCGCTCATCTTTTC 25749
1303 ..... 1303
25750 TTCTTTTATACCTGCTTTCATTTCTCTGATTTAAAGCCGATACT 25799
1304 .....Asn.GlnHisSerIleIleGluSerGlnIle 1314
25800 TCACGTGGGATCTTATCAGCAACACAGCATATATGAGCCCAAGAGATC 25849
1315 PheAsp 1316
25850 TTTCAT 25855
seq_name: gb_p14:AF057299
seq_documentation_block: 4449 bp mRNA PRI 19-AUG-1999
LOCUS AF057299
DEFINITION Homo sapiens RAD50-2 protein (RAD50) mRNA, alternatively spliced, complete cds.
ACCESSION AF057299
VERSION AF057299.1 GI:5739040
KEYWORDS
SOURCE
ORGANISM human.
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 4449)
REFERENCE Kim,K.K., Shin,B.A., Seo,K.H., Kim,P.N., Koh,J.T., Kim,J.H. and Park,B.R. Molecular cloning and characterization of splice variants of human RAD50 gene. Gene 235 (1-2), 59-67 (1999)
JOURNAL 99346140
MEDLINE 2 (bases 1 to 4449)
AUTHORS Kim,K.K., Shin,B.A., Kim,P.N., Seo,K.H., Koh,J.T. and Park,B.R. Direct Submission
JOURNAL Submitted (03-APR-1998) Pharmacology, Chonnam University, College of Dentistry, 5 Hak-Dong-Ku, Kwangju 501-190, South Korea
FEATURES
source
1..4449
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"

```



399 AsnArgIleuysAlaArg.....LeuSerSerIe 408  
: : : : :  
1513 AAACCTTGAGAGAGACAGAAAGGGAACAAAACCTCCCAACCACT 1562  
408 UGLAspLeuAspLeuAspLysLysSerAsnGluAspGluLeuAspV 425  
: : : : :  
1563 GATGATGACTTTCAGAAAAGAGACTCGAAAACAAAACAGTATAT. 1611  
425 alleuTrpLysHisTyrLeuLysIleAsnAlaArgTyrSerGluValAsp 441  
: : : : :  
1512 .....GAGTAAAGATTAAGAAACTGACCTGGGA 1641  
442 GlyIleIleGlnSerLysIleGluSerMetSerGlyIleLeuArgArg 458  
: : : : :  
1642 AGAATAATGTGATTAAATCAGAAATCTAAGT.....AA 1676  
458 GLYAspLysGluLysGluArgAspAlaAlaGluValGluLeuSerLysP 475  
: : : : :  
1677 GAAACAGAAATGAGCTGAAAATGTGAAGTATGAATTCACAGACTGGAG 1726  
475 heaenLeuSerArgIleAspGluArgGluArgHisMetGlnIleGluVal 491  
: : : : :  
1727 GATCTCAGACAGAGATCTGGAAGTACAGCAGAGCTC...ATAAAGCT 1773  
492 GLUArgLysThrLeuAlaLeuGlyGluArgAspTyrAspSerIleIleSe 508  
: : : : :  
1774 GAACTGAG.....TTAAGCAGGCTGAGAAAACAGCAATGTAGA 1814  
508 RGLIuLysArgThrGluValTyrSerLeuGluLys..... 520  
: : : : :  
1815 AACCTTAAATAAGAGATTAATAGTCTCCAAATGAAAAGCAGACTTAG 1864  
521 .....IleLysValLeuLeuArgGluLysAspIleIleAsnArgAsn 534  
: : : : :  
1865 ACAGAGACCTCTGTAACCTGACAGAGAGAGAGAGCTTAACATCAT 1914  
535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGluSe 550  
: : : : :  
1915 ACAACAACACCTACCCAAATGAGATGCTGACCAAGACAAAGCAAGCTGACA 1964  
550 rSerLysAspLysLeuAsnGluIleValAsnGluHisLysAspLysIle 567  
: : : : :  
1965 AGAT...GAAACAATTCGAAAATTAATCTAGCCACAGTATGATTA 2011  
567 yLysValLeuArgLysArgAsnProPheGluLysAspMetLysLysGlu 583  
: : : : :  
2012 CCTCAGCTTG...GGATATTTCCCAACAAAACAGCTGAGAGC... 2055  
584 IleAsnGluAlaPheTrpProValAspLysGluTyrAsnGluLeuArgSe 600  
: : : : :  
2056 .....TCG.....CTACATAG 2066  
600 rLysSerGlnGluAlaGluGlnGluLeuLysPheThrGlnSerLysValT 617  
: : : : :  
2067 TAAATCAAAAAGAAATTAATCAG..... 2088  
617 hrAspAlaArgGluGlnLeuThrLysLeuArgArgAspMet..... 630  
: : : : :  
2089 .....ACGAGGACAGACTTCCAAAATGACAAAGAACTAGCTTACT 2133  
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuG 646  
: : : : :  
2134 GAGCAGAAATAAAATCATATTAATTAATCAATAAAAGAAAGAGAGACA 2183  
646 nIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAspA 661  
: : : : :  
2184 GTTCCAGTTACGAGACAGAGCTGTTGATGTTGTGTAGCCAGGAT 2233  
661 IametaLysLysArgAspGluLysArgLysGluAsnPheAlaAsnGly 677  
: : : : :  
2234 TTGAAGATGATTAGACAGGCTTAAAGAGAAATTAAGAAAATCATCAAAA 2283

678 MetArgGluMetLeuAlaPro.....PheG 686  
: : : : :  
2284 CAGGAGCCATGTTGCTGGAGCCACAGAGCTTACCTCCAGTTCATTAC 2333  
686 uHisLeuAlaArgLysAsnHis...ValLysProCysCysGluArgAlaP 702  
: : : : :  
2334 TCAGCTAACAGACGAAACCCAGTATGTTGCCCGCTTGTCAGAGATTT 2383  
702 he...ThrProAspGluGluAspGluPheValLysLysGlnArgMetGln 717  
: : : : :  
2384 TTCAGACAGAGGCTGATCAAGAGATCACTACGTATTTGCACTTAA 2433  
718 AsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSer 734  
: : : : :  
2434 CTGCGACTTGTCTCAGATTAATCAAGTCAACAGAAATCAGAGCTTAA 2483  
734 nAlaGluAlaLeuPheGlnLeuAspLysLysLeuArgThrIleTyrAsp 751  
: : : : :  
2484 AAGGAAAGAGGCGCTGATGAATGCTGGAGCTTGCCCATGAGGCAA 2533  
751 IatYValLysLeuValGluGluThrIlePro...LeuAlaGluLys... 765  
: : : : :  
2534 GCATTAATGTATTGACAGAGAGAAATACCGAATTAAAGAAACAACTG 2583  
766 ...AsnLeuAsnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaP 781  
: : : : :  
2584 CAGATGTCAATAGACATACAGCCCTAAGAACACATAGAAAGACA 2633  
781 eAspAspLeuLeuGly..... 786  
: : : : :  
2634 AGAACAACCTGTGGTACAAATATGCTGAAAGAGAAAGTCCAAAGTAT 2683  
787 .....ValLeuAlaHisValGlnMetAspArgAspAla 797  
2684 GCCTGACAGATGTACATATATGAGAGGTTCCAGATGGAATTAAAGAT 2733  
798 ValGlu.....AlaLeuLeuGlnProThrAspTh 807  
: : : : :  
2734 GTTGAAGAAAATTTGCCAACAAGCAAGCTTAAGTACAGAAATAGAC.. 2781  
807 rIleAspArgHisValHisGluIleGlnLeuValLysGluValGluA 824  
: : : : :  
2782 .TTAGATGCAAGCTGCCAACAAGTCAACAG.....GAGAAACAG 2821  
824 sPLeuGluTyrAlaLeuAspSerSerGlyArgGlyValLysSerLeuGlu 840  
: : : : :  
2822 AGAACAAGCAACAGTACAGCA.....GTTCTAGTAAAGATT 2859  
841 GluIleGlnLeuGluLeuAsnPheLeuGlnIleArgThrArgAspThrLeu 857  
: : : : :  
2860 GAATGTGAATGCTAAGCTT..... 2877  
857 eValGluValAspAspLeuArgArgGlnHisArgMetLeuAsnGluAsp 874  
: : : : :  
2878 .....ATACAGACAGCAGAGAACAGATCAACATCTAA 2911  
874 etSerSerAlaGlnValArgTrpHisAsnAlaArgGluGluValLys 890  
: : : : :  
2912 AAATGACAAACA.....AATGAGCTTAAATCTGAAACAACTTACG 2949  
891 AlaSerSerIleLeuGluArgPheGlnLysSerGlnGluGluValLe 907  
: : : : :  
2950 ATATCCACTAATTTGCAACGTCGTCAGCAACTGGAGAGCAGACTGTGA 2999  
907 uLeuAlaGluGluLysGluGlnLeuIleValGluLysLysLeuLeuG 924  
: : : : :  
3000 ATTATCCACTAAGTCAAGTCTTTGTACAGAGAGATAAGATGCTAAG 3049  
924 IuSerLeuAspProLeu.....SerLysGluLys 933  
: : : : :  
3050 AGCAGGTAAAGCCCTTGAACAACAACTTGAAGAAATTCACCAAGAAAA 3099  
934 GluSerLeuLeuGlnGluTyrAsnAlaLeuLysGlnLysLeuAspGluG 950





BASE COUNT  
ORIGIN

2080 a 1097 c 1301 g 1414 t

```

alignment_scores:
  Quality: 1730.00
  Ratio: 1.918
  Gaps: 38
  Length: 1365
Percent Similarity: 65.126
Percent Identity: 32.419
alignment_block:
  alignment: 1
  alignment_offset: 0
  alignment_start: 0
  alignment_end: 1365
  alignment_score: 1730.00
  alignment_ratio: 1.918
  alignment_gaps: 38
  alignment_length: 1365
  alignment_percent_similarity: 65.126
  alignment_percent_identity: 32.419

```

```

alignment_block:
  US-09-538-396-2 x HSU63139      ..
Align seg 1/1 to: HSU63139 from: 1 to: 5892

```

```

1 MeSerThrValAspIserLeuIleGlyGlnIleArgSerPhe... 15
389 ATGTCCCGGATCGAANAAGATGAGCATTTCTGGCCGCGGAGTTTGGAAAT
16 AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuI 32
439 AAGAGCAAAAGAAATAAGCAAAATATACATTTCTTGACGCCCTTGACATTTT
32 LeValIleProAsnGlnLysIleGlyLysThrThrIleIleGlnCysLeuLys 48
489 TGGTGTGACCCCATGGGGGGGGAAGAGCACCATCATATGAATGCTGTAATA
49 LeuSerCysThrIleGlyLeuLeuProLysSerLysArgSerGlnIleThrPh 65
539 TATATTTGTACTGAGATTTTCCCTCCTGGAACCAAA...GGAATACATT 585
65 eValHisAspProLysValAlaIleGlyLeuThrGlnThrLysGlnIleLe 82
586 TGTACACAGATCCCAAGGTTGCTCAAGAAACAGATGTGAGCCCAAGATTC 635
82 ySerLeuArgPheLysThrIleAlaIleGlyLysAspValValCysIleArgSer 98
636 GTCTGCATTTTGCTGATGTGCATATGAGAGCAATTAATGCTGCAGAAAGATCT 685
99 PheGlnLeuThrGlnLysAlaIleSerLysMetGlnPheLysAlaIleGlnL 115
686 ATGGGTGTACTCTGAGAAAGCAAAAAGACAGAAATTTAAACCTGTGGAGG 735
115 rValIleGlnThrIleAsnProHisIleThrGlyLysIleLysLeuSer 132
736 AGTCATTT...ACGAGAACCAAGCAT...GGTGAAGAAGCTCAGCTGAGCT 779
132 yArgCysValAspMetAspArgGlnIleProIleLeuMetGlnLysIleSer 148
780 CTAAAGTGTGAGAAATTTGACGAGAAAGATGACAGTCTCTTGGGGGTTTCC 829
149 LysAlaValIleGlnLysAsnValIlePheValHisGlnAspGlnSerAsnTr 165
830 AAGGCTGTGTAATAATATGATCTATTTTCGTGATCATAGAAAGATTTTAATG 879
165 proLeuGlnAspProSerThrLeuLysLysLysPheAspIlePheS 182

```

880. GCCTTAAAGTGAAGAAAGCGCTTGGAAACCAAAAGTTGATGAGATTTT 922

182 eratIthrArgTthrLysAlaLeuGluValIleLysLysLeuHISLys 198

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

930 CAGCAAAAGATACATTTAAACCTTAGAAACACTTCGGCAGGTACGTCAG 979

199 AspGlnMetGlnGluIleLysThrPheArgLeuLysLeuGluAsnLeuGlu 215

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

980 ACACAAGCTCAAAAGTAAAGAAATATCAATTAATGGAACTAAATAATCTGAA 1029

215 nThrValLysAspGlnAlaHISLysLeuArgLysAsnIleLeuIleAspG 232

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1030 GCAATRTAAGCAAAAGCTTGCGAGATTCGCGATCAAGATTCAAGTAAG 1079

232 InGluLysSerAspAlaSerLysSerGlnMetGlnGlnLeuLysGlnLys 248

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1080 AAGCCAGATTACATCTTCAAAAGGAATTCGAAATCCTTAGAAGAAAGAA 1129

249 IleCysGlyThrGluArgGluLeuLeuLeuMetGluThrSerLeuAspGlu 265

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1130 CTTGATTCATTGAAAAACGCTAAAGAAATAATGAACTAAATCTCTGAA 1179

265 uLeuArgArgLeuGlnGlnGlnIleAspIleLysAlaThrGluArgSer 282

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1180 AATTAATGAACCTTGACAAATGAATTAAGCCTTGATGCCGAAGAAGAC 1229

282 hLeuLeuThrGlnGlnIleGluLysLeuAlaLeuSerGluGluAsn 298

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1230 AATGAGAAACATTAATAGTGAATCGAAGAGAAAGAAATGGAAGAGTTT 1279

299 GluAspThrAspGluGluLeuMetGluThrPheGlnIleLysPheGluGlu 315

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1280 CAAAGGACGATGACCAACTAAATGACTATATCAACAATCACCAAGAAC 1329

315 gIleAlaLeuLeuGluThrLysLysSerLysLeuValArgAspMetAsp 332

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1330 AGTAAGGAGAAAGAAAGAAATTTGTAGACTGTCATCGTAACTGGCAA 1379

332 SpGluAlaSerTyrSerSerValLeuSerLysGlnAsnSerGluLeuThr 348

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1380 AACTAATTAAGAACTGATGCGCTTCATCAACGAAATAATCGAACTGCT 1429

349 HISGluIleGlyLysLeuGlnAlaGluAlaAspAlaHISLeuThrMetLys 365

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1430 GTTGAACAGGGCGCTCTACAGCTGCACAAACAATGCCCATCAAGAACTAT 1479

365 sHISGluArgAspSerAspIleLysAsnIleCysThrLysHISAsnLeuG 382

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1480 CCGAGCTAGAGATTAATTAATTCACCTTTGGCAACACAGCTAGAAATTGG 1529

382. IyrProValProGluHISProPheThrAsnAspValAlaMetAsnLeuThr 398

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1530 AAGCGTTGAGCGTGACCATCATCACTGAAGACAGATTAATAAATTTTCAC 1579

399 AsnArgIleLysAlaArg.....LeuSerSerLe 408

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1580 AAACCTGTAGAGAGACACAAAGAGGGAGAGCAAAACTGCCAACCAACT 1629

408 uGluAsnAspLeuLeuAspLysLysSerAsnGlnAspGlnLeuAspV 425

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1630 GATGATGACTTTGCCAGAAAGAGACACTGAAACAAAAACAGATAGAT. 1678

425 allLeuTrpLysHISLysLeuLysIleAsnAlaArgTyrSerGluValAsp 441

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1679 .....GAGATTAAGAGATTAAGAAAACTGGACTGGGA 1708

442 GlyGlnIleGlnSerLysIleGluSerMetSerGlyIleLeuArgArg 458

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1709 AGAATATATGACTTAATAATCAAAATCTAGT.....AA 1743

458 gLysAspLysGluLysGluArgAspAlaAlaGluValGluLeuSerLysP 475

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1744 GAAGCGAATGAGCTGAAATAATGTGAAGATATGAATATCAACAGCTTGGAG 1793

```

751. lattyalllybuevalbgluuthrilepo... leuallaglyls... 765
766 ... aspleasnglnhlsleuallaspulserglnysalnglnalaph 781
767 ... 786
781 eapspasleuengly... 786
787 ... 797
788 ... 797
789 ... 797
790 ... 797
791 ... 797
792 ... 797
793 ... 797
794 ... 797
795 ... 797
796 ... 797
797 ... 797
798 ... 797
799 ... 797
800 ... 797
801 ... 797
802 ... 797
803 ... 797
804 ... 797
805 ... 797
806 ... 797
807 ... 797
808 ... 797
809 ... 797
810 ... 797
811 ... 797
812 ... 797
813 ... 797
814 ... 797
815 ... 797
816 ... 797
817 ... 797
818 ... 797
819 ... 797
820 ... 797
821 ... 797
822 ... 797
823 ... 797
824 ... 797
825 ... 797
826 ... 797
827 ... 797
828 ... 797
829 ... 797
830 ... 797
831 ... 797
832 ... 797
833 ... 797
834 ... 797
835 ... 797
836 ... 797
837 ... 797
838 ... 797
839 ... 797
840 ... 797
841 ... 797
842 ... 797
843 ... 797
844 ... 797
845 ... 797
846 ... 797
847 ... 797
848 ... 797
849 ... 797
850 ... 797
851 ... 797
852 ... 797
853 ... 797
854 ... 797
855 ... 797
856 ... 797
857 ... 797
858 ... 797
859 ... 797
860 ... 797
861 ... 797
862 ... 797
863 ... 797
864 ... 797
865 ... 797
866 ... 797
867 ... 797
868 ... 797
869 ... 797
870 ... 797
871 ... 797
872 ... 797
873 ... 797
874 ... 797
875 ... 797
876 ... 797
877 ... 797
878 ... 797
879 ... 797
880 ... 797
881 ... 797
882 ... 797
883 ... 797
884 ... 797
885 ... 797
886 ... 797
887 ... 797
888 ... 797
889 ... 797
890 ... 797
891 ... 797
892 ... 797
893 ... 797
894 ... 797
895 ... 797
896 ... 797
897 ... 797
898 ... 797
899 ... 797
900 ... 797
901 ... 797
902 ... 797
903 ... 797
904 ... 797
905 ... 797
906 ... 797
907 ... 797
908 ... 797
909 ... 797
910 ... 797
911 ... 797
912 ... 797
913 ... 797
914 ... 797
915 ... 797
916 ... 797
917 ... 797
918 ... 797
919 ... 797
920 ... 797
921 ... 797
922 ... 797
923 ... 797
924 ... 797
925 ... 797
926 ... 797
927 ... 797
928 ... 797
929 ... 797
930 ... 797
931 ... 797
932 ... 797
933 ... 797
934 ... 797
935 ... 797
936 ... 797
937 ... 797
938 ... 797
939 ... 797
940 ... 797
941 ... 797
942 ... 797
943 ... 797
944 ... 797
945 ... 797
946 ... 797
947 ... 797
948 ... 797
949 ... 797
950 ... 797
951 ... 797
952 ... 797
953 ... 797
954 ... 797
955 ... 797
956 ... 797
957 ... 797
958 ... 797
959 ... 797
960 ... 797
961 ... 797
962 ... 797
963 ... 797
964 ... 797
965 ... 797
966 ... 797
967 ... 797
968 ... 797
969 ... 797
970 ... 797
971 ... 797
972 ... 797
973 ... 797
974 ... 797
975 ... 797
976 ... 797
977 ... 797
978 ... 797
979 ... 797
980 ... 797
981 ... 797
982 ... 797
983 ... 797
984 ... 797
985 ... 797
986 ... 797
987 ... 797
988 ... 797
989 ... 797
990 ... 797
991 ... 797
992 ... 797
993 ... 797
994 ... 797
995 ... 797
996 ... 797
997 ... 797
998 ... 797
999 ... 797
1000 ... 797

```

```

3390 TGAGCTCATGACAGATATGATACAGAGATACAGAGAGAGTGG 3439
1032 TLeaspapnaLeuLysTyrArgLysThrLysAlaSerValGluGlnLe 1048
3440 CTACAGATTAACCTTACTTTAAAGAAAGAAATAGAGAACTAAAGAGT 3469
1048 uThrArgspLleGluSerLeuGluArgLeuLeuSerLleGlySerL 1065
3490 TGA.....GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3530
1065 euserAlaLleGluAlaAspLeuLysArgHisSerGlnLys..... 1079
3531 TGCAGGCTTTCAGAAATGAAAGTGAACATCAGAGTTGAGAGAGACATA 3580
1080 GluArgLeuAsnSerGluPheAsnArgTrpGlnLysThrLeuSerValTy 1096
3581 GACATATTAAGAAAGAAATCATATTTGGCATTAAGGCGACAGAAAGTTA 3630
1096 rGlnSerAsnLleSerLysHisLysGlnLysLeuLysLeuSerGlnTyrL 1113
3631 TGAAGAGAGAAATTTATCTTTTAAAGAAAGAACTTCGAGAACCAATTC 3680
1113 yAspLleGluLysArgTyrThrAsnGlnPheLeuGlnLysThrThr 1129
3681 GCGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3730
1130 GluMetAlaAsnLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAl 1146
3731 GAATCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
1146 AleuMetArgPheHisSerMetLysMetGluGlnLysLeuLysLleL 1163
3781 AATATGTGAATTTACAGATATGAAATGAAAGAAATCAATTAATATAC 3830
1163 yGlnLleTrpGlnLysThrTyrArgGlyGlnAspLleAspTyrLleSer 1179
3831 GTGACCTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3880
1180 LLeaSerAsp.....SerGlnLysAlaGlyThrArgSe 1191
3881 ATACGCTGTGATGCCGATGAAATGTATCAGCTCTGTATTAAGCGCGAA 3930
1191 rTyrSerTyrArgValValMetGlnThrGlyAspAlaGluLeuGluMet 1208
3931 TTATATACACGAGAGTGTGATGCTGAAGGAGACACACCTTGATATGC 3980
1208 rGlyArgCysSerAlaGlyGlnLysValLleuAlaSerLeuLleLeArg 1224
3981 GAGGACGATGACAGTCTGACAAAGATTAAGCTCCTCATCTCATCTGC 4030
1225 LeuAlaLeuAlaGluThrPheCysLeuAsnCySGlyLleLeuAlaLeuAs 1241
4031 CTGGCCCTGGCGTGAACGCTCTGCTCAACTGTGATCATTTGCTTGA 4080
1241 rGluProThrThrAsnLeuAspGlyProAsnAlaGluSerLeuAlaAla 1258
4081 TGAGCCACCAAAATCTTGACCGAGAAACATTAATCTCTGCACATG 4130
1258 LaleuLeuArgLleMetGluAlaArgLysGlyGlnLysAsnPheGlnLeu 1274
4131 CTCTGTGTGAGATATTAAGAGTCTGACAGACAGAGAGAGAGAGAGAG 4180
1275 LLeuAlleThrHisAspGluArgPheAlaHisLeuLleGlyGlnArgL 1291
4181 CTGGATATCATCATGATGAAGATTTGTGGAGCTTTAGGAGCTTTCGA 4230
1291 nLeuAlaGluLysTyrTyrArgValSerLysAspGluAsnGlnHisSer 1308
4231 AATGTGTGAGAAATTTCTACAGAGATTAAGAAAGACATCATCAGTCTCAG 4280
1308 LLeu 1309

```

```

4231 AGATT 4285
seq_name: gb_pat1:AR079256
seq_documentation_block: 5893 bp DNA PAT 31-AUG-2000
LOCUS AR079256
DEFINITION Sequence 44 from patent US 5965427.
ACCESSION AR079256
VERSION AR079256.1 GI:10006002
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5893)
AUTHORS Dolganov, G. and Novikov, A.
TITLE Human RAD50 gene and methods of use thereof
JOURNAL Patent: US 5965427-A 44 12-OCT-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 2081 a 1097 c 1301 g 1414 t
ORIGIN
alignment_scores:
Quality: 1730.00 Length: 1385
Ratio: 1.918 Gaps: 38
Percent Similarity: 65.126 Percent Identity: 32.419
alignment_block:
US-09-538-396-2 x AR079256
Align seg 1/1 to: AR079256 from: 1 to: 5893
1 MetSerThrValAspLysMetLeuLysGlyLleArgSerPhe.... 15
389 ATGCCCGGATCGAAAGATGATGACATCTGGCGGTGCGAGATTTGGAAT 438
16 AspProAsnLysAsnValLleThrPhePheLysProLeuThrLeu 32
439 AGAGAGCAAAAGATTAAGCAATATATCTTCTTCAAGCCCTTACAAATTT 488
32 LeuAlGlyProAsnGlyAlaGlyLysThrThrLleLleGlyCysLeuLys 48
489 TGGTTGACCAATGGCGCGGAAAGAGACATCATGATGTATAAA 538
49 LeuSerCysThrGlyLysLeuProAsnSerArgSerGlyHisThrPh 65
539 TATATTTGTACTGAGATTTCCCTCTGGAACAAA...GGAAATACATT 585
65 eValHisAspProLysValAlaGlyLysThrGluThrLysGlyGlnLle 82
586 TGTACACATCCCAAGGTGCTCAAGAAACAGATGTGAGACCCGATTC 635
82 yLeuArgPheLysThrAlaAlaGlyLysAspValValCysLleArgSer 98
636 GTCTGCATTTTCGTATCTCAATGAGAGATTAATAGCTGTCCAAAGATTC 685
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaLleGluSe 115
666 ATGGGTGTACTACGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 735
115 rValLeuGlnThrLleAsnProHisThrGlyGluLysValCysLeuSer 132
736 AGTCATT...ACTAGAACCAAGCAT...GGTGAAGAGGTGAGTGTGACT 779
132 yArgCysAlaAspMetAspArgGluLleProAlaLeuMetGlyValSer 148
780 CTAAGTGTGCAAAATTAACCGAGAAATGATCAGTCTCTTGGGGTTTCC 829
149 LysAlaValLleGluAsnValLlePheValHisGlnAspGluSerAsnTr 165
830 AAGGCTGTGTAATATATGATCTTTCTGTATCAAGAGATTTCTAATTG 879

```

```
165 pProLeuGlnAspProSerThrLeuLysLysPheAspIlePheS 182
|||||:.....:|||||:|||||:|||||:|||||:
880 GCCTTTAAAGAGAGAGGCTTTGAACGAAAGTTTGTATGATGATTTT 929
182 eAlaThrArgThrThrLysAlaLeuGluValIleLysLysLeuHisLys 198
|||||:|||||:|||||:|||||:|||||:|||||:
930 CAGCAACAAGATACATTAAAGCCTTAGAACAACCTCGCAGCTACGTACG 979
199 AspGlnMetGlnGluIleLysThrPheArgLeuLysLeuGluAsnLeuG 215
|||:|||||:|||||:|||||:|||||:|||||:
980 ACACAGAGCTCAAGAAAGTAAAGAAATGCAATGGAATCAATATATCTGA 1029
215 nThrValLysAspGlnAlaHisLysLeuArgLysAsnIleAlaGlnAsp 232
:|||||:|||||:|||||:|||||:|||||:
1030 GCAATATATAGAGAAAGCCTGTGAGATTCGTGATCAGATTACAACTAG 1079
232 GlnLysSerAspAlaSerLysSerGlnMetGlnLysGlnLysGlnLys 248
:.....:|||||:|||||:|||||:|||||:
1080 AAGCCCACTTAACATCTTCAAGGAAATGTCAAATTCCTATGACAAATGA 1129
249 IleCysGlyThrGluArgLysLeuGlnMetGlnThrSerLeuAspG 265
:.....:|||||:|||||:|||||:|||||:
1130 CTTGATCCATGACAAATCGCTCTAAAGAAATGAAATCAATCTCTCTAA 1179
265 uLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSer 282
:.....:|||||:|||||:|||||:|||||:
1180 AATATGAAGACTTGACAAATGAATTAAGCCTTGATAGCCGAAAGAAC 1229
282 hIleLeuThrGlnGlnIleHisGlnLysLeuAlaIleLeuSerGlnLys 298
:.....:|||||:|||||:|||||:|||||:
1230 AAATGAGAAAGATATATAGTGAAGTGAAGAGAAATGGAAGAAAGCTTT 1279
299 GluAspThrAspGlnGluLeuMetGlnThrGlnThrLysPheGlnGlu 315
:|||||:|||||:|||||:|||||:|||||:
1280 CAGGAGATGATGACCAACTAAATACCTATATCAACAATCACCAGAGAAC 1329
315 gIleAlaLeuLeuGlnThrLysIleSerLysLeuValArgAspMetAsp 332
:|||||:|||||:|||||:|||||:|||||:
1330 AGTAGGAGAAAGAAAGAAATGCTGACACTGCTCATCTGTAACCTGAAA 1379
332 SPGLAlaSerLysSerValLeuSerLysGlnAsnSerGlnLeuThr 348
:.....:|||||:|||||:|||||:|||||:
1380 AACTAAATAAAGAAATCTAGCGCTTCATCAACAGGAAATCAGACGCTT 1429
349 HisGlnIleGlyLysLeuGlnAlaGlnAlaAspAlaHisLeuThrMet 365
|||:|||||:|||||:|||||:|||||:
1430 GTTGAAACAGGCTCGTCTACAGCTGCACAGAGATCCGCATCAAGACAT 1479
365 sHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeu 382
:|||||:|||||:|||||:|||||:|||||:
1480 CCGAGCTAGAGATTCATTAATCAGCTTGGCAACACAGCTAGAAATGG 1529
382 LysProValProGlnHisProPheThrAsnAspValAlaMetAsnLeu 398
:.....:|||||:|||||:|||||:|||||:
1530 ATGGCTTTGAGCGTGACCATTCAGTGAAGACAGATTAATAAATTTTCA 1579
399 AsnArgGlyLeuAlaArg.....LeuSerSerLeu 408
:.....:|||||:|||||:|||||:|||||:
1580 AAACCTTGAGAGAGACAGACAAAGGCAAGCAAAATGCCCCAACCACT 1629
408 uGlnAsnAspLeuLeuAspLysLysSerAsnGlnLysAspGlnLeuAsp 425
|||:|||||:|||||:|||||:|||||:|||||:
1630 GATGATGACTTTCAGAAAGAAAGAGACCTGAAACAAAGAAACAGATGAT 1678
425 aLeuThrLysHisThrLeuLysIleAsnAlaArgThrSerGlnValAsp 441
:|||||:|||||:|||||:|||||:|||||:
1679 .....GAGATTAAGAGATTAAGAAATCGACCTGGGA 1708
442 GlyGlnIleGlnSerLysIleGlnSerMetSerGlyIleLeuArgArg 458
|||||:|||||:|||||:|||||:|||||:
1709 ACAATATATTGAGTTAAATCAGAAATCTTAAGT.....AA 1743
```

```
458 gLysAspLysGlnLysGlnArgAspAlaIleGlnValGlnLeuSerLys 475
|||||:|||||:|||||:|||||:|||||:
1744 CAGCAGAAATGAGCTGAGAAATGTAAGTATGAAATTAACAGAGTTGAAG 1793
475 hAsnLeuSerArgIleAspGlnArgGlnArgHisMetGlnIleGlnVal 491
:|||||:|||||:|||||:|||||:|||||:
1794 GATCTTCAGACGAGATCTTGAACCTGACCGAGACCTC...ATAAACT 1840
492 GluArgLysThrLeuAlaLeuGlnGlyLysArgAspThrAspSerIleIle 508
|||||:|||||:|||||:|||||:|||||:
1841 GAACTGAG.....TTAAGCAAGCTGAGAAAGAACAGCATGTAGA 1881
508 rGlnLysArgThrGlnValLysSerLeuGlnLys..... 520
:.....:|||||:|||||:|||||:|||||:
1882 AACCTTAATAAAGAGATTAATAGCTCTCCAAATGAAAGAACAGACTTAG 1931
521 .....IleLysValLeuLeuArgGlnLysAspIleIleAsnArgAsn 534
:.....:|||||:|||||:|||||:|||||:
1932 ACAGAGCCCTGCGTAACCTTGACAGAGAGATGACCAAGCAATCAATCAT 1981
535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGln 550
:|||||:|||||:|||||:|||||:|||||:
1982 ACMAACACGTAACCAATGAGATGCTGACCAAGAACAAAGCTGACAA 2031
550 rSerLysAspLysLeuAsnGlnIleValAsnGlnHisLysAspLysIle 567
:.....:|||||:|||||:|||||:|||||:
2032 ACAT...GACAAATCAGAAATTAATCTAGGACAGCTGATGAATTA 2078
567 yAlaValLeuArgGlyArgAsnProPheGlnLysAspMetLysGln 583
:|||||:|||||:|||||:|||||:|||||:
2079 CCTCACTGTG...GATATTTTCCCAACAAAGAACGCTGAAAGAC... 2122
584 IleAsnGlnAlaPheThrProValAspLysGlnLysGlnLeuArgSer 600
|||:|||||:|||||:|||||:|||||:
2123 .....TGG.....CTACATAG 2133
600 rLysSerGlnGluAlaGlnGlnLysLeuLysPheThrGlnSerLysVal 617
|||||:|||||:|||||:|||||:|||||:
2134 TAAATCAAAAGAAATTAATCAG..... 2155
617 hAspAlaArgGlnGlnLeuThrLysLeuArgArgAspMet..... 630
:|||||:|||||:|||||:|||||:|||||:
2156 ....ACGAGGACAGACTTGCCAAATGAAACAGAACTCACTCATCT 2200
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuG 646
:.....:|||||:|||||:|||||:|||||:
2201 GAGCAAAATAAATCATATTAATATGAACTAAAGAAAGAAAGAGAGA 2250
646 nIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAsp 661
:|||||:|||||:|||||:|||||:|||||:
2251 GTTGTCAGATTACGAAGACAGAGCTGTTGATGTTGTGTGACCGAGATT 2300
661 lMetAsnLysArgAspGlnLysArgLeuGlnLysAsnPheAlaAsnGly 677
:.....:|||||:|||||:|||||:|||||:
2301 TTGAAGTGAATTAGACAGCTTAAGAGAAATGAAAAAACACACAAAA 2350
678 MetArgGlnMetLeuAlaPro.....PheGln 686
:|||||:|||||:|||||:|||||:|||||:
2351 CAGCGAGCATGCTGCTGAGACACAGAGTTACTCCAGTTCACTTAC 2400
686 uHisLeuAlaArgLysAsnHis...ValCysProCysGlnLysArgAla 702
:|||||:|||||:|||||:|||||:|||||:
2401 TCAGCTAAGACAGAAACAGCATGTTGCCCCCTTTTCAGAGAGTTT 2450
702 hE...ThrProAspGlnGlnLysGlnPheValLysLysGlnArgMetGln 717
|||:|||||:|||||:|||||:|||||:|||||:
2451 TTCAGACAGAGCTGAGTACAAAGAGTCATCAGATTTGGCAGCTATAA 2500
718 AsnSerSerThrAlaGlnArgSerLysAlaLeuAlaMetGlnSerSer 734
:.....:|||||:|||||:|||||:|||||:
2501 CTCGCACTGCTCCAGATTAACATCAAGTCAACAGATCAAGCTAATAA 2550
734 nAlaGlnAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleLysAsp 751
```



1308 lelle 1309  
111  
4281 AGATT 4285

seq\_name: gb\_pat1:A86963

seq\_documentation\_block: 5893 bp DNA PAT 22-JAN-2000  
LOCUS A86963  
DEFINITION Sequence 54 from Patent WO9838306.  
ACCESSION A86963  
VERSION A86963.1 GI:6735747  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 5893)  
AUTHORS Dolganov,G.  
TITLE TRANSCRIPTS ENCODING IMMUNOMODULATORY POLYPEPTIDES  
JOURNAL Patent: WO 9838306-A 54 03-SEP-1998;  
GENELABS TECH INC (US)

FEATURES  
source 1..5893  
/organism="unidentified"  
/isolate="RAD50\_SEQ"  
/db\_xref="taxon:32644"  
389..4327  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CA669369.1"  
/db\_xref="GI:6735748"  
/translation="MSRIEKMSILGVRSGIEDKDKQITFFSPRLTIVGPNAGKTT  
IIECLRYICTGDPPTGKNTFVHDKVQAQETDVRAQIRLQFHDVNGELIAYGRSMVC  
TQSKTEKTEKTEGVTTRTHGKGVLSLSKCAIDEMISLSIDVSKLVNNTVEJHCE  
DSWMPISDEGKALOKFDEIFSAATRYIKALETPOVOTOGOKVEYMEYKIKOYE  
KACEIRPDQTSKFAOLTSSEKYEKYEENETDPLKKNLKEIENHLSKIMKIDNEKALD  
SKRQMEKDNSELEKMEKFOCTDEQNDLNNHQTVEKEREKLYDCRRELEKAK  
ESRLNKEKSELVEQGRLOLQDRHOEHTRANDSLQSLATQLELDGFERGPSEKQ  
IKNFHLVROEGEAKTANQNLNDEPAKETTLOKQIDELDKTGTGRLIEKSEL  
SKRONLKNVYELQLESGDRILELQDELIAERLSKAEKSNVETLKMEYLSIQ  
NEKADLDRLTKLQDMEQNLHHTTTRTOMEMLTKQKADDEQIRIKSHSDELTL  
LGFPNKKOLEDLHSKSEKINOTRRLAKLNLKELASPNKHNINNELKRRREOLSS  
YEDKLTQVGSQDFESDLRLKEIEKSKQRMAGATVYSOFTIQLTDENOSCP  
VCCRVPTQTEALEQVLSLQSLRLAPDKLSTSELEKKEKREDEMLGLVPMQSTII  
DLKEKIPELRNKLVNVRDILQKNDIEQOETLGIPEEESAKVCLDVIMERF  
OMELKVEREIAQAAKQIDIDRLTQVQNOKEQKQELDTVSSKTELRKLIQOQ  
OEOIOLKSTTNELKSEKLOISTNLOROOLEQOTVELSTGSLYEIRIDAKVOYSP  
LETTLEKPOQEKELINKRTSKIKADKNDIKKRYKNHGVNKDIENTIQGCKNDY  
KKQKTELKNVIALQLECEKHEKINEDMKLMQDIDTQRIQERWLDNLTLEKRNNE  
LKEVEERKQHLKEMQOMVLOKMSHOKLEENIDNTRNNHNLALGRQKYEETIHF  
KKELEPERFADAEKREMMIVKRTTELNVKDDIYKTLQDAIMKFSMKMEIINI  
IRDLMSYRQGDIEYIEIRSDADENYASDKRRNRYRVMLKGDPLALMRGSCSAG  
OKVLASLIIRLALAEFFELCNGIILADEPTDNDRENIESLAHLVETIISRSQQRNF  
QLLVYTHDEPFVLLGRSEVEKFRKKNIDOCSELYKCSVSSLGNVH"

CDS

BASE COUNT 2078 a 1098 c 1303 g 1414 t  
ORIGIN

alignment\_scores:

Quality: 1729.00 Length: 1385  
Ratio: 1.919 Gaps: 38  
Percent Similarity: 65.054 Percent Identity: 32.419

alignment\_block:  
US-09-538-396-2 x A86963 ..

Align seg 1/1 to: A86963 from: 1 to: 5893

1 MetSerThrValAspLysMetLeuIleuLysGlyIleArgSerPhe..... 15  
||||| :||||| ||| :|||||  
389 ATGTCTCCGAGTCGAAAGATGACATCTCTGCGCGCGAGATTGGAGAT 438  
16 AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuI 32

439 AGAGCAAAAGATTAAGCAATATACATCTCTGACCCCTTACAAATT 488  
432 leValGlyProAsnGlyAlaGlyLysThrThrIleIleGlyLysLeuLys 48  
489 TGGTTGGACCAATGGCGGGAAGAGACACCATCATTTGAATGCTAA 538  
49 LeuSerCysThrGlyLeuLeuProProAsnSerArgSerGlyHisThrPh 65  
539 TATATTCTACTCGAGATTCCTCTCTGAAACCAAA...GGAAATACATT 585  
65 eValHisAspProLysValAlaGlyLysThrGlyThrLysGlyLysIleL 82  
586 TGTACACGATCCCAAGTGTCTCAAGAAACAGATGTGAGACCCGATTC 635  
82 yLeuArgPheLysThrAlaGlyLysAspValValCysIleArgSer 98  
636 GCTGCAATTCGTGATGTCAATGAGAACTATAGCTGTGCAAGATCT 685  
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSe 115  
666 ATGGTGTGTACTCAGAAAAGCAAAAGACAGAAATTTAAACTCTGGAAG 735  
115 rValLeuGlnThrIleAsnProHisThrGlyLysValCysLeuSer 132  
736 AGTCATT...ACTAGAACAAAGCAT...GCTGAAAGGTGACGTGACCT 779  
132 yArgCysAlaAspMetAspArgGluIleProAlaLeuMetGlyLysIse 148  
780 CTAACTGTGCAAGAAATGACCGAGAAATGATCAGTCTTGCGGGTTC 829  
149 LysAlaValLeuGluAsnValIlePheValHisGlnAspGluSerAsnTr 165  
830 AAGGCTGGCTAAATAAATGCTATTTTCTGCTACAGAGAACTCAATG 879  
165 ProLeuGlnAspProSerThrLeuLysLysAspPheAspIlePheS 182  
880 GCTTTAAGTGAAGAAAGGCTTGAAGCAAAAGTTTATGAGATTTT 929  
182 eValThrArgThrThrLysAlaLeuGluValIleLysLysLeuHisLys 198  
930 CAGCAACAAGATACATTAAGCTTGAAGACCTCGGAGGTAGCTCAG 979  
199 AspGlnMetGlnIleLysThrPheArgGlyLysLeuGluAsnLeuG 215  
980 ACACAAAGGCAAGAAAGTAAAGAAATATCAAAAGGACCTAAATATCTAA 1029  
215 nThrValLysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAsp 232  
1030 GCAATATTAAGAAAGGCTGTGAGATTCGTGATCAGATTACAAATAG 1079  
232 GlnGlyLysSerAspAlaSerLysSerGlnMetGlnLysGlyLys 248  
1080 AAGCCACGATTAACATCTTCAAAAGAAATGTCAATTCATAGAAATGA 1129  
249 IleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAsp 265  
1130 CTGTGATTCATTAAGAAATGCTTAAAGAAATGAATGAATCACTCTCA 1179  
265 uLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSer 282  
1180 AATTAATGAACCTTGACMAGNATTAAGCCCTTGATGATCCCAAGAAC 1229  
282 hrLeuLeuThrGlnGlnHisGlyLysLeuAlaAlaLeuSerGlnGluAsn 298  
1230 AATGAGAGAAAGATATAAGTGAAGTGAAGAAATGAAGAAAGTTT 1279  
299 GluAspThrAspGlnGluLeuMetGluThrGlnThrLysPheGlnGluArg 315  
1280 CAAGGACTGATGACCACTAATGATCATATATCAATCAACAGAGAAC 1329  
315 gIleAlaLeuLeuGluThrLysIleSerLysLeuValArgAspMetAsp 332

```
1330 AGTAAGGAGAAAGAAAGAAATGTAGACTGTATCGTGAACCTGGAA 1379
332 spgluAlaSerTyrSerValLeuSerIysGlnAnSerGluLeuThr 348
1380 AACATAATTAAGAAATCTAGGCTTCTCAATCAGGAAATATCAGAACTGCTT 1429
349 HisGluLeuGlyLeuGlnAlaGluAlaAspAlaHisLeuThrMetIly 365
1430 GTTGAACAGGGTCTCTACAGCTCCAAAGCAGATGCCATCAAGAACAATAT 1479
365 HisGluArgAspSerAspIleLeuAsnIleCysThrLysHisAnLeuG 382
1480 CCGAGCTAGAGATTCATTAATTCAGTCTTGGCAACACAGCTAGAAATGG 1529
382 LysProValProGlnHisProPheThrAsnAspValAlaMetAspLeuThr 398
1530 ATGGCTTGGACCTGGGCCATTCAGTGAAGACAGATTAATAATTTTCAC 1579
399 AsnArgIleLysAlaArg.....LeuSerSerLe 408
1580 AAATCTGTGAGAGAGACAGCAAGAGGGAACCAAACTGCCAACCACT 1629
408 uGluAsnAspLeuAspLysLysLysSerAsnGluAspGlnLeuAspV 425
1630 GATGAATGAGACTTGCAGAAAGAGACTGTGAACCAAAACAGATAGAT. 1678
425 allLeuTyrPheHisTyrLeuLysIleAsnAlaArgTyrSerGluValAsp 441
1679 .....GAGATAAGAGATTAAGAAACTGCACTGGGA 1708
442 GlyGlnIleGlnSerLysIleGlnSerMetSerGlyIleLeuArgArg 458
1709 AGAATAATTCGTTAAATCAGAAATCTTAAGT.....AA 1743
458 gLysAspLysGluLysGluArgAspAlaAlaGluValGlnLeuSerLysP 475
1744 GAACAGAGATGAGCTGAAGAAATGTGAAGTATGATACAGACAGTGAAG 1793
475 heaAnLeuSerArgIleAspGlnArgGlnArgHisMetGlnIleGluVal 491
1794 GATCTTCAGACAGATTCCTGAACTGACAGCAGAGCTC...ATTAAGCT 1840
492 GluArgLysThrIleuAlaLeuGlyGluArgAspTyrAspSerIleIleSe 508
1841 GAACGTGAG.....TTAAGCAAGCTGACAGAAACACCAATGTAGA 1881
508 rGlnLysArgThrGluValTyrSerLeuGlnLys..... 520
1882 AACCTTAATAATGAGATTAATAGTCTCAAAATGAAAGAACAGACTTAG 1931
521 .....IleLysValLeuLeuArgGluLysAspIleIleAsnArgAsn 534
1932 ACAGAGACCTGCTGAACCTTGACAGAGATGAGAGCTTAACCAACAT 1981
535 AlaAspGluArgValLysLeuGly...LeuLysArgAspAlaLeuGlnSe 550
1982 ACAACAAACACCTACCCAAATGAGAGATGCTGACCAAGCAACAAAGCTGACA 2031
550 rSerLysAspLysLeuAsnGlnIleValAsnGlnHisLysAspLysIleL 567
2032 AGAT...GAACAAATCAGAAATTAATCTAGGCACACAGTGAATTA 2078
567 yLysValLeuArgGlyArgAsnProPheGluLysAspMetLysLysGlu 583
2079 CCTCAGCTGTTG...GGATATTTTCCCAACAAACAAAGCTTGAAGAC... 2122
584 IleAsnGlnAlaPheTyrProValAspLysGluLysArgSerGlnLeuArgSe 600
2123 .....TGG.....CTACATAG 2133
600 rLysSerGlnGluAlaGlnGlnGlnLysLysPheThrGlnSerLysValT 617
2134 TAATCAAAAGAAATTAATCAG..... 2155

617 hisAspAlaArgGluGlnLeuThrLysLeuArgArgAspMet..... 630
2156 ....ACCAGGACAGACTTGCCAAATTTGAACAGAACTAGCTTCACT 2200
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuG 646
2201 GAGCAGATTAATAATCAATATTAATTAATGAACTAAAGAAAGAAAGACA 2250
646 nIleSerAlaAsnValAspMetPheProLysValLeu....GlnAsp 661
2251 GTTGTCCAGATTACGAAACAAAGCTGTGTGATTTGTGTAGCCAGAT 2300
661 IamLeuAsnLysArgAspGlnGlnLysArgLeuGlnAsnPheAlaAsnGly 677
2301 TTGAAAGTGAATTTAGACAGGCTTAAGAGAAATTTGAATAATCATCAAAA 2350
678 MetArgGluMetLeuAlaPro.....PheG 686
2351 CAGGAGCCATGCTGCTGGAGCCACAGACTTACTCCAGTTCATTAC 2400
686 uHisLeuAlaArgLysAsnHis...ValCysProCysCysGluArgAlaP 702
2401 TCAGCTAACAGACGAAACAGATCATGTGCTGCTGCTGCTGCTGCTGCTG 2450
702 he...ThrProAspGluGluAspGluPheValLysLysGlnArgMetGln 717
2451 TTCAGACAGAGGCTGAGTTCACAAAGAGTCAATGATTCAGCTTAA 2500
718 AsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSerAs 734
2501 CTGCGAGCTTGCTCCAGATTAACCTCAAGTCAACAGATCAGAGCTTAA 2550
734 nAlaGluAlaLeuPheGlnGlnLeuAspLysLysLeuArgThrIleTyrAsp 751
2551 AAAGGAAAGGCGGCTGATGAATGCTGGAGCTGCTGCCCTGAGGCAAA 2600
751 lArgValLysLeuValGlnGluThrIlePro...LeuAlaGluLys... 765
2601 GCATTAATGATTTGAGAGAGAGAAATACCAATTAATGAGAAACAACTG 2650
766 ...AsnLeuAsnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaP 781
2651 CAGAAATTCATAGACATACAGAGCTTAAGAAACGACATAGAAAGACA 2700
781 eAspAspLeuLeuGly..... 786
2701 AGAAACACTTGGGTACAATTAATGCTGAAGAAAGAGTCCAAAGTAT 2750
787 .....ValLeuAlaHisValGlnMetLysArgAspAla 797
2751 GCCTGACAGATGTACATTAATGAGAGGTTCCAGATGGAACCTTAAGAT 2800
798 ValGlu.....AlaLeuGlnGlnProThrAspTh 807
2801 GTTAAAGAAATAATTCGCAACAAAGCAGCTAGCTACAGAAATAGAC... 2848
807 rIleAspArgHisValHisGluIleGlnGlnLeuValLysGluValGlu 824
2849 TTAGATCGAACTGTCACACAGTCAACAG...GAGAAACAG 2888
824 sPheLeuLysValAlaLeuAspSerSerLysArgGlyValLysSerLeuGln 840
2889 AGAAACAGCAACAGTGAACCA.....GTTTCTAGTGAAT 2926
841 GluIleGlnLeuGlnLeuAsnPheLeuGlnArgThrArgAspThrLeuI 857
2927 GAATGAAATCTAGCTT..... 2944
857 eValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGluAspM 874
2945 .....ATACAGACAGCAGAGAACAGATTCACATCTAA 2978
```



```

8831 GTACCTGTGGCAGACTACCTATCGTGGACAGATTTGAATTCATAGAA 3880
1180 IleasnSerasp.....SergLgLYlaGLyTharGse 1191
||||| ||| ||| |||
8881 ATACGGCTGATCCGATGAAATGTATATGATCGCTTGATTAAGGCGGAA 3930
1191 rTySerTyraValValMetclInThrGlyaspAlaGluLeuGluMetA 1208
||||| ||| ||| |||
9393 TTTAACTACCGCAGTGTGTATCTGTGAAGGAGACACAGCCTTGGAATATC 3980
1208 rglGlyarGySserAlaGlyGlnLYValLeuAlaSerLeuIleIeary 1224
||||| ||| ||| |||
3981 GAGAGCATGACAGTGTGGACAAAGGATATTAGCTTCATCATCATTCGC 4030
1225 LeuAlaLeuAlaGluThrPhacYSLeuAsnCYasGlyYlLeuAlaLeuAs 1241
||||| ||| ||| |||
4031 CTGGCCCTGGCTGAACGTTCTGCTCTAATGTGGATCTATTCGCTGGA 4080
1241 pGluProThrThraSLeuAspGlyProAsnAlaGluSerLeuAlaAla 1258
||||| ||| ||| |||
4081 TGACCAACAACAATCTTGACCGAGAAACATGTGATCTTGACATG 4130
1258 lAlaLeuAlaGlylIeMetGlnAlaArgLYsgLYngInGluAsnPhgInLeu 1274
||||| ||| ||| |||
4131 CTCTGGTGTGAGATATATAAAAGTGTCTCAACAGCAGCCTAATCTTCAGCT 4180
1275 lIleValIleThrHISAspGluArgPhaAlaHISLeuIleGlyGlnArg 1291
||||| ||| ||| |||
4181 CTGTTAATCACTCATGATGAAGATTTTGTGAGCTTTTAGAGACTGTGA 4230
1291 nLeuAlaGluLYsTyTyTyraValSerLYsAspGluAsnGlnHISser 1308
||||| ||| ||| |||
4231 ATATGTGGAGAAATTTCTACAGATTAATAAAGAACATCATCATGATGCTCAG 4280
1308 lIle 1309
|||||
4281 AGATT 4285
name: gb_pat1:AR048375
documentation_block: 5893 bp DNA PAT 29-SEP-1995
US AR048375 Sequence 54 from patent US 5821091.
INSTRON AR048375
AR048375.1 GI:5970718
ORFS
ORF1 Unknown.
ORF2 Unknown.
ORF3 Unknown.
ORF4 Unknown.
ORF5 Unknown.
ORF6 Unknown.
ORF7 Unknown.
ORF8 Unknown.
ORF9 Unknown.
ORF10 Unknown.
ORF11 Unknown.
ORF12 Unknown.
ORF13 Unknown.
ORF14 Unknown.
ORF15 Unknown.
ORF16 Unknown.
ORF17 Unknown.
ORF18 Unknown.
ORF19 Unknown.
ORF20 Unknown.
ORF21 Unknown.
ORF22 Unknown.
ORF23 Unknown.
ORF24 Unknown.
ORF25 Unknown.
ORF26 Unknown.
ORF27 Unknown.
ORF28 Unknown.
ORF29 Unknown.
ORF30 Unknown.
ORF31 Unknown.
ORF32 Unknown.
ORF33 Unknown.
ORF34 Unknown.
ORF35 Unknown.
ORF36 Unknown.
ORF37 Unknown.
ORF38 Unknown.
ORF39 Unknown.
ORF40 Unknown.
ORF41 Unknown.
ORF42 Unknown.
ORF43 Unknown.
ORF44 Unknown.
ORF45 Unknown.
ORF46 Unknown.
ORF47 Unknown.
ORF48 Unknown.
ORF49 Unknown.
ORF50 Unknown.
ORF51 Unknown.
ORF52 Unknown.
ORF53 Unknown.
ORF54 Unknown.
ORF55 Unknown.
ORF56 Unknown.
ORF57 Unknown.
ORF58 Unknown.
ORF59 Unknown.
ORF60 Unknown.
ORF61 Unknown.
ORF62 Unknown.
ORF63 Unknown.
ORF64 Unknown.
ORF65 Unknown.
ORF66 Unknown.
ORF67 Unknown.
ORF68 Unknown.
ORF69 Unknown.
ORF70 Unknown.
ORF71 Unknown.
ORF72 Unknown.
ORF73 Unknown.
ORF74 Unknown.
ORF75 Unknown.
ORF76 Unknown.
ORF77 Unknown.
ORF78 Unknown.
ORF79 Unknown.
ORF80 Unknown.
ORF81 Unknown.
ORF82 Unknown.
ORF83 Unknown.
ORF84 Unknown.
ORF85 Unknown.
ORF86 Unknown.
ORF87 Unknown.
ORF88 Unknown.
ORF89 Unknown.
ORF90 Unknown.
ORF91 Unknown.
ORF92 Unknown.
ORF93 Unknown.
ORF94 Unknown.
ORF95 Unknown.
ORF96 Unknown.
ORF97 Unknown.
ORF98 Unknown.
ORF99 Unknown.
ORF100 Unknown.
ORF101 Unknown.
ORF102 Unknown.
ORF103 Unknown.
ORF104 Unknown.
ORF105 Unknown.
ORF106 Unknown.
ORF107 Unknown.
ORF108 Unknown.
ORF109 Unknown.
ORF110 Unknown.
ORF111 Unknown.
ORF112 Unknown.
ORF113 Unknown.
ORF114 Unknown.
ORF115 Unknown.
ORF116 Unknown.
ORF117 Unknown.
ORF118 Unknown.
ORF119 Unknown.
ORF120 Unknown.
ORF121 Unknown.
ORF122 Unknown.
ORF123 Unknown.
ORF124 Unknown.
ORF125 Unknown.
ORF126 Unknown.
ORF127 Unknown.
ORF128 Unknown.
ORF129 Unknown.
ORF130 Unknown.
ORF131 Unknown.
ORF132 Unknown.
ORF133 Unknown.
ORF134 Unknown.
ORF135 Unknown.
ORF136 Unknown.
ORF137 Unknown.
ORF138 Unknown.
ORF139 Unknown.
ORF140 Unknown.
ORF141 Unknown.
ORF142 Unknown.
ORF143 Unknown.
ORF144 Unknown.
ORF145 Unknown.
ORF146 Unknown.
ORF147 Unknown.
ORF148 Unknown.
ORF149 Unknown.
ORF150 Unknown.
ORF151 Unknown.
ORF152 Unknown.
ORF153 Unknown.
ORF154 Unknown.
ORF155 Unknown.
ORF156 Unknown.
ORF157 Unknown.
ORF158 Unknown.
ORF159 Unknown.
ORF160 Unknown.
ORF161 Unknown.
ORF162 Unknown.
ORF163 Unknown.
ORF164 Unknown.
ORF165 Unknown.
ORF166 Unknown.
ORF167 Unknown.
ORF168 Unknown.
ORF169 Unknown.
ORF170 Unknown.
ORF171 Unknown.
ORF172 Unknown.
ORF173 Unknown.
ORF174 Unknown.
ORF175 Unknown.
ORF176 Unknown.
ORF177 Unknown.
ORF178 Unknown.
ORF179 Unknown.
ORF180 Unknown.
ORF181 Unknown.
ORF182 Unknown.
ORF183 Unknown.
ORF184 Unknown.
ORF185 Unknown.
ORF186 Unknown.
ORF187 Unknown.
ORF188 Unknown.
ORF189 Unknown.
ORF190 Unknown.
ORF191 Unknown.
ORF192 Unknown.
ORF193 Unknown.
ORF194 Unknown.
ORF195 Unknown.
ORF196 Unknown.
ORF197 Unknown.
ORF198 Unknown.
ORF199 Unknown.
ORF200 Unknown.
ORF201 Unknown.
ORF202 Unknown.
ORF203 Unknown.
ORF204 Unknown.
ORF205 Unknown.
ORF206 Unknown.
ORF207 Unknown.
ORF208 Unknown.
ORF209 Unknown.
ORF210 Unknown.
ORF211 Unknown.
ORF212 Unknown.
ORF213 Unknown.
ORF214 Unknown.
ORF215 Unknown.
ORF216 Unknown.
ORF217 Unknown.
ORF218 Unknown.
ORF219 Unknown.
ORF220 Unknown.
ORF221 Unknown.
ORF222 Unknown.
ORF223 Unknown.
ORF224 Unknown.
ORF225 Unknown.
ORF226 Unknown.
ORF227 Unknown.
ORF228 Unknown.
ORF229 Unknown.
ORF230 Unknown.
ORF231 Unknown.
ORF232 Unknown.
ORF233 Unknown.
ORF234 Unknown.
ORF235 Unknown.
ORF236 Unknown.
ORF237 Unknown.
ORF238 Unknown.
ORF239 Unknown.
ORF240 Unknown.
ORF241 Unknown.
ORF242 Unknown.
ORF243 Unknown.
ORF244 Unknown.
ORF245 Unknown.
ORF246 Unknown.
ORF247 Unknown.
ORF248 Unknown.
ORF249 Unknown.
ORF250 Unknown.
ORF251 Unknown.
ORF252 Unknown.
ORF253 Unknown.
ORF254 Unknown.
ORF255 Unknown.
ORF256 Unknown.
ORF257 Unknown.
ORF258 Unknown.
ORF259 Unknown.
ORF260 Unknown.
ORF261 Unknown.
ORF262 Unknown.
ORF263 Unknown.
ORF264 Unknown.
ORF265 Unknown.
ORF266 Unknown.
ORF267 Unknown.
ORF268 Unknown.
ORF269 Unknown.
ORF270 Unknown.
ORF271 Unknown.
ORF272 Unknown.
ORF273 Unknown.
ORF274 Unknown.
ORF275 Unknown.
ORF276 Unknown.
ORF277 Unknown.
ORF278 Unknown.
ORF279 Unknown.
ORF280 Unknown.
ORF281 Unknown.
ORF282 Unknown.
ORF283 Unknown.
ORF284 Unknown.
ORF285 Unknown.
ORF286 Unknown.
ORF287 Unknown.
ORF288 Unknown.
ORF289 Unknown.
ORF290 Unknown.
ORF291 Unknown.
ORF292 Unknown.
ORF293 Unknown.
ORF294 Unknown.
ORF295 Unknown.
ORF296 Unknown.
ORF297 Unknown.
ORF298 Unknown.
ORF299 Unknown.
ORF300 Unknown.
ORF301 Unknown.
ORF302 Unknown.
ORF303 Unknown.
ORF304 Unknown.
ORF305 Unknown.
ORF306 Unknown.
ORF307 Unknown.
ORF308 Unknown.
ORF309 Unknown.
ORF310 Unknown.
ORF311 Unknown.
ORF312 Unknown.
ORF313 Unknown.
ORF314 Unknown.
ORF315 Unknown.
ORF316 Unknown.
ORF317 Unknown.
ORF318 Unknown.
ORF319 Unknown.
ORF320 Unknown.
ORF321 Unknown.
ORF322 Unknown.
ORF323 Unknown.
ORF324 Unknown.
ORF325 Unknown.
ORF326 Unknown.
ORF327 Unknown.
ORF328 Unknown.
ORF329 Unknown.
ORF330 Unknown.
ORF331 Unknown.
ORF332 Unknown.
ORF333 Unknown.
ORF334 Unknown.
ORF335 Unknown.
ORF336 Unknown.
ORF337 Unknown.
ORF338 Unknown.
ORF339 Unknown.
ORF340 Unknown.
ORF341 Unknown.
ORF342 Unknown.
ORF343 Unknown.
ORF344 Unknown.
ORF345 Unknown.
ORF346 Unknown.
ORF347 Unknown.
ORF348 Unknown.
ORF349 Unknown.
ORF350 Unknown.
ORF351 Unknown.
ORF352 Unknown.
ORF353 Unknown.
ORF354 Unknown.
ORF355 Unknown.
ORF356 Unknown.
ORF357 Unknown.
ORF358 Unknown.
ORF359 Unknown.
ORF360 Unknown.
ORF361 Unknown.
ORF362 Unknown.
ORF363 Unknown.
ORF364 Unknown.
ORF365 Unknown.
ORF366 Unknown.
ORF367 Unknown.
ORF368 Unknown.
ORF369 Unknown.
ORF370 Unknown.
ORF371 Unknown.
ORF372 Unknown.
ORF373 Unknown.
ORF374 Unknown.
ORF375 Unknown.
ORF376 Unknown.
ORF377 Unknown.
ORF378 Unknown.
ORF379 Unknown.
ORF380 Unknown.
ORF381 Unknown.
ORF382 Unknown.
ORF383 Unknown.
ORF384 Unknown.
ORF385 Unknown.
ORF386 Unknown.
ORF387 Unknown.
ORF388 Unknown.
ORF389 Unknown.
ORF390 Unknown.
ORF391 Unknown.
ORF392 Unknown.
ORF393 Unknown.
ORF394 Unknown.
ORF395 Unknown.
ORF396 Unknown.
ORF397 Unknown.
ORF398 Unknown.
ORF399 Unknown.
ORF400 Unknown.
ORF401 Unknown.
ORF402 Unknown.
ORF403 Unknown.
ORF404 Unknown.
ORF405 Unknown.
ORF406 Unknown.
ORF407 Unknown.
ORF408 Unknown.
ORF409 Unknown.
ORF410 Unknown.
ORF411 Unknown.
ORF412 Unknown.
ORF413 Unknown.
ORF414 Unknown.
ORF415 Unknown.
ORF416 Unknown.
ORF417 Unknown.
ORF418 Unknown.
ORF419 Unknown.
ORF420 Unknown.
ORF421 Unknown.
ORF422 Unknown.
ORF423 Unknown.
OR
```

389 ATGCCCGATCGAAAGATGAGCATCTGGGCGTGGGAGTTTGGAT 438  
16 .AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeu 32  
AGAGGACAAAGATAGCAAAATATACATTCTTCCAGCCCTTCAAAATTT 488  
32 LeValGlyProAsnGlyValAlaGlyLysThrThrIleIleGlyLeuLys 48  
TGGTTGGACCCCAATGGGGGGGAGAGACACCATTCATTAATGCTTAA 538  
49 LeuSerCysThrGlyGlyLeuLeuProProAsnSerArgSerGlyHisThrPh 65  
TATATTTGCTACTGAGATTTCCCTCCGGAACCAA...GGAATFACAT 585  
65 eValHisAspProLysValAlaGlyGlyThrGlyThrLysGlyLeuIle 82  
TGTACACAGATCCCAAGGTTGCTCAAGAAACAGATGTAGACCCAGATTC 635  
82 yLsLeuArgPheLysThrAlaAlaGlyLysAspValValCysIleArgSer 98  
GTCGTCAATTTTGTGATGTCATGAGAGAACTTATAGCTGTGCAAGATCT 685  
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGlyLeu 115  
ATGGTGTCTACTCAGAAAAGCAAAAAGCAGAAATTTAAACTCTGGAAG 735  
115 rValLeuGlnThrIleAsnProHisThrGlyGlyLysValCysLeuSer 132  
AGTCAT...ACTAGAACAAAGCAT...GGTGAAGAGTCACTGAGCT 779  
132 yArgCysAlaAspMetAspArgGluIleProAlaLeuMetGlyValSer 148  
CTAAGTGTGCAAAATGATCCGAGAAAGATCACTTCTTGGGGTTC 829  
149 LysAlaValLeuGluAsnValIlePheValHisGlnAspGlnSerAsnTr 165  
AAGGCTGTCTACTCAGAAAAGCAAAAAGCAGAAATTTAAACTCTGGAAG 879  
165 pProLeuGlnAspProSerThrLeuLysLysLysPheAspIlePheS 182  
GCCTTTAACTGAAGAAAGCTTGAAGCAAAAGTTGATGATTTT 929  
182 eAlaThrArgTyThrLysAlaLeuGluValIleLysLysLeuHisLys 198  
GAGCAAGATACATTAAGCTTAGAACCACTCGGACGTCAGTCAG 979  
199 AspGlnMetGlnGluIleLysThrPheArgLeuLysLeuGluAsnLeuG 215  
ACACAAAGCTGAGAAAGTAAAGATATCAATTCGAACATAATATCTGAA 1029  
215 nThrValLysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAsp 232  
GCAATATATGAGAAAGCTGTGAGATCTGTCAGATTCAGATTAAGTAAG 1079  
232 InGlyLysSerAspAlaSerLysSerGlnMetGluGlnLeuLysGlyLys 248  
AAGCCCACTTACATCTTCAAGAGAAATGTCAATCTCATAGAGATGAA 1129  
249 IleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspG 265  
CTGATTCATGAGAAAGATCTTAAAGAAATTAACATATCTCTCTAA 1179  
265 uLeuArgTyArgLeuGlnGlyIleLysIleAspIleLysAlaThrGluArgSer 282  
AATATGAACTGACATGAAATTAAGCTTGAATGCGAAGAAC 1229  
1180 hLeuLeuThrGlnGlnHisGlyLysLeuAlaAlaLeuSerGlnGluAsn 298  
AAATGAGAGAAAGATTAATGTAAGTGAAGAAAGAAAGGATTTT 1279  
1230 gLysAspThrAspGluGluLeuMetGluTrpGlnThrLysPheGlnGluArg 315  
CAAGGACTGATGAGCAACTAAATGACTTATATCAATTCACAGAGAAC 1329

315 gIleAlaLeuLeuGluThrLysIleSerLysLeuValArgAspMetAsp 332  
AGTAAAGAGAGAAAGAAAGAAATGTTAGACTGTCATCTGACTGGAAA 1379  
332 spGluAlaSerTySerSerValLeuSerLysGlnAsnSerGluLeuThr 348  
AATTAATTAAGATCTAGGCTTCTCAATCAGAAATAATGAACTGTT 1429  
349 HisGluIleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThrMetLys 365  
GTTGACAGGGTCTCTACACCTGCAAGCAGATCCCATCAAGAACATAT 1479  
365 sHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeuG 382  
CCGAGTACAGATTCATTAATTAATGCTTTGGCAACACAGCTAAATGG 1529  
382 LyrProValProGluHisProPheThrAsnAspValAlaMetAsnLeuThr 398  
ATGGCTTTGAGCTGGGCCATTCAGTGAAGACAGATTAATAATTTTCAC 1579  
399 AsnArgIleLysAlaArg.....LeuSerSerLe 408  
AACTGTGAGAGAGAGACAGAGAGGAGGAGCAAAACTGCCAACCACT 1629  
408 uGluAsnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLeuAspV 425  
GATGATGACTTTTCAGAAAGAGACTCTGAACCAAAACAGATAGT 1678  
425 alleuTrpLysHisTyLeuLysIleAsnAlaArgTySerGluValAsp 441  
GAGTAAAGATTAAGAAACTGGACTGGGA 1708  
442 GlyGlnIleGlnSerLysIleGluSerMetSerGlyIleLeuArgArgAr 458  
AGATTAATGATTAATAATCAGAAATCTAGT.....AA 1743  
458 gLysAspLysGlyLysGluArgAspAlaAlaGluValGluLeuSerLysP 475  
GAACAGAGATAGCTGAAATAATGTAAGTATTAATGACAGTTGGAG 1793  
475 hAsnLeuSerArgIleAspGluArgGluArgHisMetGlnIleGluVal 491  
GATCTTCAGACAGATTTCTTGACTGACACAGAGCTC...ATAAACCT 1840  
492 gLysArgTyThrLeuAlaLeuGlyLysArgAspTyArgSerIleIleS 508  
GAAGCTGAG.....TTAAGCAAGCTGAGAAACACCATGTGAA 1881  
508 rGlnLysArgTyThrGluValTyrSerLeuGlnLys..... 520  
AACCTTAAATGAAGATTAATTAAGTCTCAAAATGAAGAACACAGACTTGA 1931  
521 .....IleLysValLeuLeuArgGluLysAspIleIleAsnArgAsn 534  
ACAGAGACCTGCTGAACCTTACCAAGAGATGAGAGCACTTAACCATCAT 1961  
535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGlyLeu 550  
ACAACACACGCTTACCAAAATGAGATGCTGACCAAGCAAGCAAGTGAACA 2031  
550 rSerLysAspLysLeuAsnGluIleValAsnGlnHisLysAspLysIle 567  
AGAT...GAAACAAATCAGAAATAATTAATCTAGCAGACAGTATTAATTA 2078  
567 yLysValLeuArgGlyArgAsnProPheGluLysAspMetLysGlu 583  
CTTACCTGTTG...GGATATTTCCCAACAAACAAACGCTGAAGAC... 2122  
584 IleAsnGlnAlaPheTrpProValAspLysGluTyArgAsnLeuArgS 600  
TGG.....TGG.....CTACATAG 2133



```

3781 AATATGAATTCACAGTATGAAATGAGAAATCAATATAATATAC 3830
1163 ysglulntutpglmglnthrtttrarglglmlnspileasptylliser 1179
      ::::::::::::::::::::::::::::::::::::::::::::::
3831 GTGACCTGTGGCAAGTACTATCGTGACAAATATGATGATACATGAA 3880
1180 llaasnerasp..... SerGlUGLYAlaGlylthrtargse 1191
      ||| ::|
3881 ATACGCTCTATGCCATGAAATGATATCAGCTTCTGATTAAGCGCGGA 3930
1191 ttrsertrargvalvalmetglnthrglaspalagluLeuGluMetA 1208
      ::::::::::::::::::::::::::::::::::::::::::::::
3931 TTTAACTACGAGTGTGTATGCTGAAGGAGACACAGCCTTGATATGC 3980
1208 rglgltargcysserallaglylglulysvalleuAaserleullearg 1224
      ||| ::|
3981 GAGGACATCATGCTGCGGACAAAGATATTAGCTCCTACATCATTCGC 4030
1225 leuAlaLeuAlaGluThrPheCysLeuAsnCysGlylleuLeuAlaLeuAs 1241
      ||| ::|
4031 CTGGCCCTGCTGAAACGTTCTGCTCACTGCTGCGCTATTCCTTGGA 4080
1241 pgluprotthtrAsnleuaspglproasnalagluSerleuAlaAla 1258
      ||| ::|
4081 TGAGCCACAACAATCTTGACGAGAAACATTGATCTCTTGCAATG 4130
1258 laleuLeuargllemetgluAlaarglylglulnluasnphleuLeu 1274
      ||| ::|
4131 CTCTGTTGGATTAATTAATAAGTCGTCACGACGCTAATCTTCAGCTT 4180
1275 lilevallethrhlshspgluargpheaAlaAlaIleuIleGlylnArGg 1291
      ||| ::|
4181 CTGGTATATCTCATGATGAGATTTTGTGGAGCTTTTGAAGAGCTTGA 4230
1291 nleuAlaGluLysTyrTyrArgValserLysAspGluAsnGlnHisert 1308
      ||| ::|
4231 ATATGTGAGAAATTCACAGATTAATAAGAACATCATGATGCTGACAG 4280
1308 lelle 1309
      |||
4281 AGATT 4285

seq_name: gp-pr9:HSRAD50
seq_documentation_block: 4123 bp mRNA PRI 11-DEC-1997
LOCUS HSRAD50 DEFINITION H.sapiens mRNA for RAD50.
ACCESSION 275311 VERSION 275311.1 GI:2687852
KEYWORDS RAD50.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4123)
AUTHORS Offenberg, H.H.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4123)
AUTHORS Offenberg, H.H.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1996) Offenberg H.H., Agricultural University,
Genetics, Wageningen 2, Wageningen, 6703 HA The Netherlands
FEATURES
Source
1..4123
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="testis"
/clone_lib="library HL161a in lambda gt10 (Clontech)"
43..3999
/gene="RAD50"
43..3999
/gene="RAD50"
/function="DNA repair and recombination protein"

```

```

/codon_start=1
/product="RAD50 homologue hsrAD50"
/db_xref="GI:2687853"
/db_xref="GI:2687853"
/translation="MLIFSVRDMFAKMSILVRSFSDKDKQITTFSPPLTILVGN
GAGKTTIECKLYICTGDPPTKGTNGTVDKQVADVDVRAQILQFRODVGELIIV
ORSMVCPKQSKKTEKTELEVTTRKHGKXSLSKAEIDREMISSISGVKAVANNV
IFCHQEDSNWPLSESKALKOKFDEFTSRTIKALETTRVQROTQGVETOMELKY
LKQYKACELIDQITTSKAEQTSSELYSENLDPKRNLEIENHLSKIMKLDN
EIKALDSRRKQOMENKSELEKMEKVEFGOTBOLNDLHNRQVREKRLVCHRE
LEKLNKESRLNOEKSELVROGRQLOADRHQEHIRADSLISLATOLEDGERG
PESEROINPFRHVEBOGEAKTANOLMOPAEKTEIKOKOIDEIRPKKGLGRIIE
LKSLELKKQNEKVKVKELOOBESSRIIELOELIKARELSKAKNSVETLKA
EVLSONEKADIDRLKLDQMEDLNHTTTRQMEHLTDKADQDQIRIKSRHS
DELTSILQYFPNKKOLEDMLSKSEIKINQTDRLKLNKLEASQNNHINNELER
EEOUSSYEDKLFEDVAGSODFSDLRLEEKLESKQAMAGATAVASFOTYOTDE
NOSCCPQOVRFQVEAELOEALISDQSKRLAPDKLSTESLKKKEKRDQMLGLAP
MROSTIIDKEKEIPELRNKLQNVNDIQRNDIEOETLIGTIMPEESA KVCITLV
TIMERFQMLKDVETKINQAKKLOGIDLDITVOQVONKQKOKLDTVSKIELN
KLIDQOQOIOHLKSTNELSEKLQISTNQROQLEOTVELSTEVOVSLYRETKDA
KEOVSPLLETLEKFOQKEELINKNTSNKIAODKLANDIKKVNINHGYMDIENHIO
DGKDYMKOKETELNKVIAOLSECEKHEKINEDKRLMODIDOKIOERMLQDNLTJ
RKRNBELEKVEBEGKQHLKEMGOMVOLMKSEHOKLEENIDINKRNHNLALGRQYE
EELIHFKELREPOPRDAEEKYREMIYMTTELYNKRDLIYITLDQAIKPFHMKM
EINKITIDLRSTYRGODIEYIEKRSADENVSASDKRNINTRYVWKDQDALDKR
GRCSAGQKVLASLIIRLALFTFLNCGIILADEPTNLDEENIESLHAIVETIKSR
SQORNFQLVVTHEDFVPELLGRSRYVKKFYRIKNDQCEIYKCVSLSGFNVH"

BASE COUNT 1630 a 661 c 901 g 931 t
ORIGIN

alignment_scores:
Quality: 1711.00 Length: 1375
Ratio: 1.899 Gaps: 35
Percent Similarity: 65.527 Percent Identity: 31.855

alignment_block:
US-09-538-396-2 x HSRAD50
Align seg 1/1 to: HSRAD50 from: 1 to: 4123

6 LysMetLeuIleLysGlyIleArgSerPhe..... AspProAspAsnLys 20
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
76 AAGATGAGCATTTCTGGCGTGGCGAGATTGTGAAATGAGGACAAAGTAA 125
20 saenValIleThrPhePheLysProLeuThrLeuIleValGlyProAsnG 37
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 GCAATTATACCTTCTCTGAGCCCTTACAAATTTGCTGGAGCCCATG 175
37 lyaIaGlyLysThrIleIleGluCysLeuLysLeuSerCysThrGly 53
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
176 GGCGGGGAAAGACGACCATCATGTAATGTCTAAATATATTATTACTGCA 225
54 GluLeuProProAsnSerArgSerGlyHisThrPheAlaHisProLys 70
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 GATTTCCTCTCTGGAACCAAA...GGAAATACATTGTGTACACGATCCCA 272
70 svaIaIaGlyLysThrGluThrLysGlyGlnIleLysLeuArgPheLys 87
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
273 GGTTCCTCAAGAAACAGATGTGAGCCCGCATTCGCTGCATTTTCGTC 322
87 hvalaIaGlyLysAspValValCysIleArgSerPheGlnLeuThrGln 103
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
323 ATGTCAATGAGAACTTAATAGCTGCAAAAGATCATGCTGCTACTACAG 372
104 LysAlaSerLysMetGluPheLysAlaIleGluSerValLeuGlnThrI 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
373 AAAAGCAAAAGACAGAAATTAAGCCCTGGAAGAGATCTT...ACTAG 419
120 eaenProHisThrGlyLysValCysLeuSerTyrArgCysAlaAsp 137
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 AACAAAGCAT...GTGAAAGGTCAGTCTGACCTCAAGTGTGCAAGAA 466

```

430 yLeuLysIleAsnAlaArgYrSerGIuValAspGIuInIleGlnSer 446  
:::|||||  
1351 ..... GAGTATGAACATAAGAAACTGCATCGGGAGAAGATTATTCAGCTTA 1395

447 LysIleGluSerMetSerGIyIleuNrgArgArgGLysAspLySGIuLy 463  
||| ||| ::|||  
1396 AAATCAGAAATCCCTAAGT..... AAGAAGCAAGATTAGAGCT 1430

463 sGIuArgAspAlaAlaGIuValGIuLeuSerLYPheAsnIeuSerArgI 480  
:::::  
1431 GA AAAATGGAAGTAGTGAATTATACACAGCACTGTGAAGAGATCTTCACACAGA 1480

480 IeAspGIuArgGIuArgHisMetGlnIleGIuValGIuArgLyThLeu 496  
:::|||||  
1481 TTTCTTGAACTGGACACGAGAGCTC... ATAAAAGCTGAACGTGAG.... 1521

497 AlaleuGIuLArgAspTYrAspSerIleIleSerGlnTyArgTrnGI 513  
||||| :::  
1522 ... TTAAAGCAAGCTGACGAAAAACAAGCAATTAGAACCTTAAAAATGGA 1568

513 uValTYrSerIeuGIuInTyS..... IleYsv 523  
||| | |||||  
1569 AGTAATAGTCTCCAAAATGAAAAAGCAGACTTGAACAGCACCTGGGTA 1618

523 AlIeuLeuArgGIuLyAspIleIleAsnArgAsnAlaAspGIuArgVal 539  
||| ::||| :  
1619 AACCTTGACCAGGAGTAGTGAGAGCTTAAACCATCATACAAACACGTAAC 1668

540 LysLeuGIy... LeuLYLysAspAlaIeuGIuSerLYsAspLYle 555  
:::|||||  
1669 CAATGTGAGATGCTGACCAAAGACAAACACTGCAAAAGT... GANCAAT 1715

555 uAsnGIuIleValaGIuHisLYsAspLYleLYsLYsValIleuArg 572  
:::||||| :  
1716 CAGAAAATAAATAATCAGGACAGAGTAGAATTAACTCACTGTTG... G 1762

572 LYArgAsnProPheGIuLYsAspMetLYsGIuInIleAsnGlnAlaPhe 588  
:::|||||  
1763 GATATTTTCCC..... AACCAAAAAACACTTGAAGACCTGGCTA 1800

589 TPProValAspLYsGIuTYrAsnGIuLeuArgSerLYsSerGIuInVal 605  
:::|||||  
1801 CATAGTAAATCAAAGAAATTTATATGACACAGGACACACTTGCCAAT 1850

605 agLIuGIuIleuLYsPheThrGlnSerLYsValYThrAspAlaArgGIu 622  
:::::|||||  
1851 GAACAAGACAACTA..... GCTTCATCTGAGCAGA 1879

622 InIeuThrLYsLeuArgArgAspMetAspAlaLYsArgArgPheLeuAsp 638  
::: :::  
1880 ATPAAATCATATTAATATATGAACTAGAAAGAAAGAGACAGACTGTGCC 1929

639 SerLYsLeuGlnSerIleuGIuInIleSerAlaAsnValaAspMetPhePr 655  
:::|||||  
1930 AGTTACGAAGACAAAGCTGTTATATGTGTGGTACCAAGATTTTGAAG 1979

655 oLysVal..... LeuGlnAspAlaMetAsnLYsArgAspGIuInLYsa 670  
:::||||| :  
1980 TGATTTAGACAGGCTTAAAGAGAAATTTGAAAAATCAATCAAAACAGCGAG 2029

670 rGIeugIuAsnPheAlaAsnGIuMetArgGIuMetLeuAlaProPheGIu 686  
:::|||||  
2030 CCATGCTGGCTGAGACACAGACAGTTTACTCCCAATTC..... ATTACT 2073

687 HisLeuAlaArgLYsAsnHis... ValGlySProCYsGIuArgAlaPh 702  
:::::|||||  
2074 CAGCTTACAGACGAAACCAAGCATGTGGCCCCGTTTTTCAGAGAGTTT 2123

702 e... ThrProAspGIuGluAspGIuPheValLYsLYsGlnArgMetGlnA 718  
||| ||| ::|||  
2124 TCAGACAGAGAGCGTGAATCTCAAGAAAGCCATCGATTTTGCAGGTGTAAAC 2173

718 asnSerThrAlaGIuArgSerLYsAlaLeuAlaMetGluSerSerAsn 734



[illegible]

934 AAGCCAGTTAGCTCTTCGACGAAATGTCAAGTCTCAAGAGATAG 983  
249 ILeysglYhrgluArgluIleuGlmetCuhthrsrLeuaspGl 265  
984 CTTCAGCCATTGAAGAAATCGCTGAAAGAGATTGAACAACTGTCTAA 1033  
265 uLeuArgArpLeuGlIngluIleAspIleYsAlaThrGluArgSert 282  
1034 AATATGAACCTTATATGAATTAATTAAGCTTGAGACCGCAAGAAAC 1083  
282 hLeuLeuThrgIngluIngluYsLeuAlaLeuSerGluLuan 298  
1084 AAATGAGAAAGATTAACGTAATTAGACGAAATGAATAAGTTT 1133  
299 GluAspThrsArgluIleuMetGluThrgInthrsrPhegluArg 315  
1134 CAAAGGACTGTAGACGCTAAATGACTGTATCAATCAATCCAGCAAC 1183  
315 gIleAlaLeuLeuGluThrsIleSerLysLeuValArgAspMetAsp 332  
1184 AGTAAGGAGAAAGAAAGCCCTGTAGACTGCCAGCAGAACTGGAGA 1233  
332 spGluAlaSerTySerSerValLeuSerLysGluAsnSerGluLeuThr 348  
1234 AGTGAATTAAGAGACGCGCTCTCAACCAAGAAAGACAGACTGCTT 1283  
349 hIsGluIleGluYsLeuGluAlaGluAlaAspAlaHisLeuThrMet 365  
1284 GTGAGAGAGGCTGTACAGCTACAGCAGCAGATCCATCAAGACACAT 1333  
365 sHIsGluArgAspSerAspIleYsAsnIleCyThrLysHisAsnLeu 382  
1334 CCGAGCCAGACACTATTAATTCAGTCTTGCGCAACACTTGAATGG 1383  
382 LysProValProGluHisProPheThrsAspValAlaMetAsnLeuThr 398  
1384 ATGGTTTGAAGCTGACATTCAGTGAAGACAGATTAAATTTTCAT 1433  
399 AsnArgIleYsAlaArg.....LeuSerSerLe 408  
1434 GAACCTTGAAGAGACAGAAAGAAAGCTAAATGCCACCCAGCT 1483  
408 uGluAsnAspLeuAspLysLysLysSerAsnGluAspGluLeuAsp 425  
1484 CTTCAGGACCTGACACAAAGAAAGCTTGAACAGACAGCAGCTTAT 1532  
425 alleuThrsHisTyLeuLysIleAsnAlaArgTySerGluValAsp 441  
1533 .....GAGCTGAGGACAGAGAGAGAGGAGCTGGG 1562  
442 GluGluInleGlnSerLysIleGlnSerMetSerGlyIleLeuArgArg 458  
1563 AGGAGAGTTGAGCTGAGAGACAGAGATCTCAG.....AA 1597  
458 gLysAspLysGluYsArgAspAlaIleGluValGluLeuSerLysP 475  
1598 GAACCAAGACGAGCTGCGGACAGTGAAGAGTGAAGCTCAGACAGTGA 1647  
475 heAsnLeuSerArgIleAspGluArgIleGluHisMetGluIleGluVal 491  
1648 GCTCCTCCAGAGATCTGGAACCTGAGACAGAGCTC.....ACGAAAGCT 1694  
492 GluArgLysThrLeuAlaLeuGluYsLysArgAspTyAspSerIleIle 508  
1695 GAAGGTAA.....TTAAGCAAGCCGAGAAATAGACACATAGA 1735  
508 rGluLysArgThrGluValTySer.....LeuG 518  
1736 AACCTTAAGCAAGAGATTAAGAGCTCCAAATGAGAAAGCTGACCTGG 1785  
518 luGluLysIleYsValLeuLeuArgIleYsAspIleIleAsnArgAsn 534  
1786 ACAGAGCTTGCGAAAGCTGAGACAGAGATGAGCAGTTAAACATCAT 1835

535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGluSe 550  
1836 ACAACAAACCCGACACAGATGAGATGCTCACCAAGACAAACTGACAA 1885  
550 rSerLysAspLysLeuAsnGluIleValAsnGluHisLysAspLysIle 567  
1886 AGAT...GAACAGATCAAGAAATTAAGTCTAGGACAGATGATGACTAA 1932  
567 yLysValLeuArgGlyArgAsnProPheGluLysAspMetLysGlu 583  
1933 CTTCAGCTGTG...GGGTACTTCTCT.....AACCAAAAAACAG 1967  
584 IleAsnGluAlaPheThrProValAspLysGluThrsArgLeuAspGse 600  
1968 CTTCAGAGCTGGCTTCATTTAAATCCAAAGAAATTAATGAGCAGCAGGA 2017  
600 rLysSerGluGluAlaGluGluGluLeuLysPheThrGlnSerLysValT 617  
2018 CAGACTTGCCAAACTGAACAAAGAGCTA.....G 2046  
617 hAspAlaArgGluGluLeuThrsLysLeuArgArgAspMetAspAlaLys 633  
2047 CTTCAGCTGAACAAATTAATTAATGATTAATGAGCTTAAGAAAGAAAG 2096  
634 ArgArgPheLeuAspSerLysLeuGlnSerIleLeuGlnIleSerAlaAs 650  
2097 GAAGAACAGTTGCTCAGTTATGAAGATTAAGCTATTTATGTTGTGTAG 2146  
650 nValAspMetPheProLysVal.....LeuGluAspAlaMetAsnLysA 665  
2147 CCAAGATTTGGAAGATGATTTGGGACAGCTTAAGAAAGAAATTAAGAA 2196  
665 rGAspGluGluArgLysArgLeuAsnPheAlaAsnGlyMetArgLumet 681  
2197 CTTCAAGACGAGCGAGCTGCTGCGACAGACCCACAGAGATTACTCCAG 2246  
682 LeuAlaProPheGluHisLeuAlaArgLysAsnHis...ValCysProC 697  
2247 TTC.....ATTACGAGCTGACAGATGAAGAAACGAGCTGCTGCCAGT 2290  
697 sCySGluArgAlaPhe...ThrProAspGluGluAspGluPheValLysL 713  
2291 CTTCACAGAGATTTACAGACAGAGCTGAGTTACAGAAAGTCTACAGCG 2340  
713 ySgInArgMetGluAsnSerSerThrAlaGluArgSerLysAlaLeuAla 729  
2341 ACTTGACGTCAAGCTGAGCTGCTCCAGATAAATCAAGTCAACAGAA 2390  
730 MetGluSerSerAsnAlaGluAlaLeuPheGluGlnLeuAspLysLeu 746  
2391 TCGGAAGCTGAAGAAAGAGAGCGGCGCTGATGAATAATGCTGGGCTGT 2440  
746 gThrIleTyAspAlaTyValLysLeuValGluGluThrIleProLeuA 763  
2441 GCCCGTAGGCAAGCATTAATGATTTGAAGAAAGAAATACCAAT 2490  
763 lAgLysAsnLeuAsnGlnHisLeuAlaAspGluSerGluLysAlaGln 779  
2491 TAAAGAACCACTACAGAGTGTCAATAGACATACAGCCCTTAAGAT 2540  
780 AlaPheAspAspLeuLeuGluValLeuAlaHisValGlnMetAspArg 796  
2541 GACATAGAACAGCAGGAACACCTTGGTGAACATATGCTGCTGAGAGAG 2590  
796 AlaValAlaLeuLeuGluInProThrAspThrIleAspArgHisValH 813  
2591 AAGTCTAAAGTGTGCTGACAGATGTTACATCATGAGAGGTTCCAGA 2640  
813 lsgLulieGlnLeuValLysGluValGluAspLeuGluTyAlaLeu 829  
2641 TGGAGCTGAAGAGATGTGAAGAGAAATTCACACAGCAGCAGCTAACCTG 2690





**JOURNAL**  
Submitted (22-DEC-1999) Biochemistry and Molecular Biology,  
Louisiana State University Health Sciences Center, 1901 Perdido  
St., New Orleans, LA 70112, USA

## FEATURES

## Source

1. .4444

/organism="Rattus norvegicus"  
/db="ncf" + version="10116#

```

/db_xref="taxon:10116"
155 4004

```

CDS

```
/function="DNA double-strand break repair and
```

recombination"

```
function="telomere length maintenance"
```

/note="similar to *Saccharomyces cerevisiae* Rad50; null mutation lethal in murine embryonic stem cells"

```
mutation is letn
/codon start=1
```

```

/coach_start=1
/product="Bad50"

```

```

/protein_id="AAFG91229.1"
/product="Kad50-

```

```

/procell_10=AAE91223
/adb_xref="GT:9651650"

```

```

/uu_xie-61:5051050
/translation="MSRIEKMSTLGVRSGIEDPKDOIISFSPPLTVGPNGAKTT

```

BASE COUNT	1549 a	840 c	1116 g	939 t
ORIGIN				

```
alignment_scores:
  Quality: 1686.00      Length: 1404
  Ratio: 1.871          Gaps: 37
  Percent Similarity: 64.174  Percent Identity: 31.197
```

```
alignment_block:
```

Align seg 1/1 to: AF218576 from: 1 to: 4444

```

1  MeSerThrValAspArgLeuIleLysGlyIleArgSerPhe..... 15
156 ATGTCCCGGATCCAAAGATGAGCACTCTGGCGCGGCAATTTTGGGAT
16  AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuI 32
206 AGAGGATTTAAAGATTAACCAATATATCTTTCTTCAGCCCTCCACAAATTT
32  LeValGlyProAsnGlyValAsnLysThrThrIleIleGluCysLeuLys 48
256 TGGTGGACCCCAATGGGGGGGGGAAAGACAGCACCATATGAAATGCTAAG
49  LeuSerCysThrGlyLysLeuProProAsnSerArgSerGlyHisThrPhe 65
306 TAAATTTGTACTGAGATATTTCCCTCCTGGAAACCAAA...GGAAATACATT 352
65  eValHisAspProLysValAlaIleGlyIleThrGluThrArgSerLysIleLeu 82
353 TGGTTCATGATCCCAAGGTTGGCTCAAGAAACAGATGGGTGCCCAAAATTC 402

```

82	ysleuicrphelyshthrtaalaaglytysaspaValcyslletrser	98
83	.....	99
403	gcctcagatttgcatgtcATGAGAGAGATGTACTTGTGCAGAGGTCC	452
404	.....	403
453	atgccttgacgacagaaaaagTAAAAAACCCGAATTTAAACCCCTGGAAG	502
454	.....	415
415	ttaleuuglnthrTleasnprohIsthGlygylusvalcysleusert	132
416	.....	503
503	atcattactagaaTAAAG.....CACGGTAAAAAGCACTGCTCAGCT	546
504	.....	132
132	ytarcyslaaspmelaspargluilepcolaleuemetgylalser	148
133	.....	547
547	ccaaTGTGTCaAAATTCGACCGAGAAATGATAAAGTTGCTTGGGGTTTCC	596
548	.....	149
149	lysalaValleuglnasnullphevalhIsglnaspluserAsntr	165
150	.....	597
597	AGTGTGTGCTTAAACAATGTTTCTTCTCCACCAAGAGACTCAAAATTG	646
598	.....	165
165	proleuglnaspproserthleuLysLysLyspheaspaSptlepes	182
166	.....	647
647	gcctcTAASTGAGGAAGAGGCTCTGAGCGAATTTGATGGAATTTT	696
648	.....	182
182	eralathtatgtTthrTlysalaLeugluVallleLysLysLeuhIlys	198
183	.....	697
697	CAGCAACAAGTATCAATTAAGCCCTAGATACGCTTCGACAGGTACGACAG	746
698	.....	199
199	asplmetgIngluileLysThrRheargleuLysleuglnasnuLeuGl	215
200	.....	747
747	ACACAAGGTCaAAAGTAAAMAATGTCAACAGATGTGAATTAATCTGAG	796
748	.....	215
215	nthrVallysaspglnaIhIstLysleuargLusnIle.....	228
216	.....	797
797	gcAAATTAAGGAGAAAGCTTGAGATCCGAAATCAGACTAGACTAGTAGG	846
798	.....	229
229	.....AlaglnaspgIngluLysSeraspalaserLyssergin	241
230	.....	847
847	AAGCCAGATTACATCTTCACGCGAAATTTGTCAAAAGCCATAGAGATGAG	896
848	.....	242
242	metgIngluInleuLysIleCysglYthrgluargluileleuGl	258
243	.....	897
897	CTTGTGACCTTAAGAAATCGC.....CTGAAGAA	925
898	.....	258
258	metgInthrSerleuAspGluLeuargrgeuIngluInIleaspi	275
259	.....	926
926	gATTGACATAACCTCTCTAAATATATAGACTTGACAATGAATTAAG	975
927	.....	275
275	leLysalaThrgluargserThrLeuLeuThrgInglInIsgluLysleu	291
276	.....	976
976	CTTGTGATAGCAAGAAAGCAAAATCGAAAAAGATAAACGTGAATTAGAA	1022
977	.....	292
292	AlaaleuSerGluInuasnIuAspThrAspGluInleuemetgIthr	308
293	.....	1026
1026	CAGAAGATCGAAAGAGTTTTCAAAGGACTGTAGCAGCACTAAAGACTT	1077
1027	.....	308
308	pclInthrLysphegIngluArgIleAlaleuLeuInuThrLysIleSerL	325
309	.....	1076
1076	GTATCAATATCAACCCAGAGAACTGTGAAGGAGAAAGAGCCCTGTGTAG	1122
1077	.....	325
325	ysleuValhIrgaspmelaspargluIhIstLysSerSerValleuser	341
326	.....	1126
1126	ACTGTACGCGTAACGTGGAAGAGCTGAGTAAAGAAAGCTCGGCTCTCAAC	1177
1127	.....	342
342	lysGlnAsnSerGluLeuThrIstgluilegIlyLysleuGlnAlaIsglnaI	358
343	.....	1176
1176	CAGGAAGAGCAGACTCTGTTGTGAGCAGAGGCTGCTATCAACATCAGAGC	1222
1177	.....	356
356	aspaIhIstleuThrmetLysIsgluLysaspserspIleLysasnI	375
357	.....	1226
1226	AGATGACATTCAGAAGCATATCCGAGCCAGAGATCACTGATTAAGTCTT	1275



2857 TGACAGAGAAATAAGAGTCTTAAGACCAATTAACCTTTGGAGATA 2906  
 929 .....LeuSerLySGluLySerLeuLeuGlnGluTyrAs 941  
 2907 GCACCTGGAAAAGTTGCAGACAGAAAAGAAATGATCCACACAAAAA 2956  
 941 nAlaLeuLySGlnLyLeuAspGluGluTyrHisGlnLeuAlaGluTyr 958  
 2957 TACAGATACAAAATGGCTCGAGTAAGATCATATCATCAAGAAAG 3006  
 958 yArGluPheGlnGlnGlnLeuAspAlaLeuGluTyrGluAsnMetLyS 974  
 3007 TTAATAATATCATGCTTACATG.....AAAGAC 3035  
 975 IleLySGluTyrLeuAspSerLyLeuAsnGlu.....LysLeuLySGl 989  
 3036 ATAAGATATTTATTTCAAGATGAAAAGATGACTATTAAGAAAGAAA 3085  
 989 uLeuGlnGluTyrHisValLeuLySHisSerGlnLeuGlnSerCysMet 1006  
 3086 AACTGAACCTTAATGAGTTGTCATT.....CAACTAATGATGATGACA 3129  
 1006 lAlLySGlnGlnArgIleSerAlaGluLeuAsnLySserLySGluLeu 1032  
 3130 AACCAAGAAAGATTAATAAGAAATGGAGAACCATGAGCAAGATATT 3179  
 1023 GlnGlnGlnGlnGlnLeuLySArgAsnIleAspAsnLeuLySTyrAr 1039  
 3180 GACACGAGAAAGATACAGAAAGTGTGTACAGATTAACCTTACACTGAG 3229  
 1039 gLySerThrLySAlaAspValGlnGlnLeuThrArgAspIleGluSerLeu 1056  
 3230 AAAAAGAGAGAGAGACTTAAGAAAGTTGAA.....GAGACAGCA 3270  
 1056 lGluTyrGluLeuSerIleGlySerLeuSerAlaIleGluAlaAspLeu 1072  
 3271 AGCAACATTTAAGAGAGATGGAGATGCAAGTTTACAGATGAAA... 3317  
 1073 LysArgHisSerGlnGlnLySGluArgLeuAsnSer..... 1084  
 3318 AATGAGACCCAGAGTTGGAAGACATTAACACGATTAAGAGAAATCA 3367  
 1085 .....GluPheAsnArgTyrGlnGlnGlnLeuSerValTyrGlnSer 1099  
 3368 CAGTTGGCATTTGGCGCACAGAAAGC.....TATCAAGAG 3405  
 1099 snIleSerLySHisLySGlnGlnLeuLySLeuSerGlnTyrLyAspIle 1115  
 3406 AAATCTTCACTTAAGAGAGAGCTCGAGAACTCAGTTCCGGAGTCA 3455  
 1116 GluLySArgTyrThrAsnGlnPheLeuGlnLeuLySThrThrGluMetAl 1132  
 3456 GAGAGAGAGTACAGAGAAATGATGATGTGATGAGAACCAAGAGCTGTG 3505  
 1132 aAsnLySAspLeuAspArgTyrTyrThrAlaLeuAspLySAlaLeuMet 1149  
 3506 GAATTAAGACCTGGACATCTTACTACAAACCTTTCATCAGCAATTAATGA 3555  
 1149 rGpHisSerMetLySmetGlnGlnLeuLySLeuLySLeuLySLeu 1165  
 3556 AATTTCACAGATGAAATGAGAAATCAACAAATTAATTCGTGATCTT 3605  
 1166 TyrPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1182  
 3606 TGGCGGAGTACCTATCGTGGCAAGATTTGAATACATAGAAATTCATC 3655  
 1182 rAsp.....SerGlnGlnAlaGlnGlnGlnGlnGlnGlnGlnGln 1194  
 3656 CGATGCTGAGAAATATATCATGCTTCTGATTAAGCGCAATTAATCACT 3705  
 1194 yTyrGlnValMetGlnThrGlyAspAlaGluLeuGluMetArgGlyArg 1210  
 3706 ACCGAGTGTGATGCTGAGAGGGGATACAGCCTTGACATGCGGGAGCA 3755

1211 CysSerAlaGlnGlnLySValLeuAlaSerLeuIleIleArgLeuAla 1227  
 3756 TGCACTGCTGGACAGAAAGTCTGCTCTCTCATCATCCAGCTGGCCTT 3805  
 1227 uAlaGluThrPheCysLeuAsnGlyIleLeuAlaLeuAspGluProt 1244  
 3806 GCGGAAACCTTGTGTGAGACTGGCGCATCTTCCTTGATGAGCCTA 3855  
 1244 hThrAsnLeuAspGlyProAsnAlaGluSerLeuAlaAlaLeuLeu 1260  
 3856 CAACAACCTGGACCGAGAAACATCGAGTCTGCGCACATGCTTGCTT 3905  
 1261 ArgIleMetGluAlaArgLySGluGlnGlnAsnGlnGlnGlnGln 1277  
 3906 GAGATTTTAAAAAGTCTGCGACAGCGCACTTCACGCTTGCTGTCAT 3955  
 1277 eThrHisAspGluArgPheAlaHisLeuIleGlyGlnArgGlnLeuAla 1294  
 3956 CACTCAGCATAGATTTTGTGAGCTCTCTAGACAGATCTGATATGCG 4005  
 1294 lLysTyrTyrArgValSerLySAspGluAsnGlnHisSerIleGlu 1310  
 4006 AGAATTTCTACAGAGTGGAAGAAACATCGACAGTGTCTCAGAGATTGTC 4055  
 1311 SerGlnGlnIle 1314  
 4056 AAGACAGCATC 4067  
 seq\_name: gb\_Pf4:AF057300  
 seq\_documentation block:  
 LOCUS AF057300 4384 bp mRNA PRI 19-AUG-1999  
 DEFINITION Homo sapiens truncated RAD50 protein (RAD50) mRNA, complete cds.  
 ACCESSION AF057300  
 VERSION AF057300.1 GI:5739042  
 KEYWORDS  
 SOURCE  
 ORGANISM human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 4384)  
 Kim,K.K., Shin,B.A., Seo,K.H., Kim,P.N., Koh,J.T., Kim,J.H. and  
 Park,B.R.  
 Molecular cloning and characterization of splice variants of human  
 RAD50 gene  
 Gene 235 (1-2), 59-67 (1999)  
 JOURNAL  
 MEDLINE 99346140  
 REFERENCE 2 (bases 1 to 4384)  
 AUTHORS Kim,K.K., Shin,B.A., Kim,P.N., Seo,K.H., Koh,J.T. and Park,B.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-APR-1998) Pharmacology, Chonnam University, College  
 of Dentistry, 5 Hak-Dong Dong-Ku, Kwangju 501-190, South Korea  
 FEATURES  
 source  
 1. 4384  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /map="5q31"  
 /cell\_line="Hela"  
 1. 4384  
 /gene="RAD50"  
 /note="DNA recombination repair gene"  
 674. 4195  
 /gene="RAD50"  
 /note="RAD50-3: a 65-bp fragment compared to RAD50-2 is  
 deleted by alternative splicing"  
 /codon\_start=1  
 /product="truncated RAD50 protein"  
 /protein\_id="RAD50326.1"  
 /db\_xref="GI:5739043"  
 /translation="MISSING:KAVLVNVIFFCHQDSNPLSEGRALKQKPEIRSAT  
 RYKMLETRQVRQYOGQVKKEYQVHEKLYKKKACIRIQYITSKEAQLSSKEIV

KSYENELDLKRLKEIEHNTSKIMKLDNEIKALSDSKKOMKONSELEKMEKVTG  
 TDEOLNDLYHNHRTVEREKRLVDCHRELEKLNKSRLLNOKSELLVQGRLOJA  
 DRHOEHIRARDSLIOSLAIOLEDFERGFSPROKINPKILVREOREGAKRANOM  
 NDBAEKETILAKOKIDIDETROKKTGLRIIEKSEILSKQNEKLVNRYELOOLGSSNR  
 ILELDDELIAKRELELSKAERNSEVETLMEVLSLONKADLDTRKLEDOEMOLNH  
 TTRTQMEMTIDKADKDEQIRIKSRHSDLSLGLFENKQLEDMHKSKEINO  
 TRRLAKLNKELASEQONKHNINNELKREKESQEDKFEVCGSDPFSDRLKE  
 EIKSSKORAMLAGATAVVSQFITOLDEMOSCPVORFOTEALEOEIYDLOSFL  
 RLAPDLKSTRESLEKREKREDEMGLVPMOSITIDLEKEITELNKLONVNDTOR  
 LKNDIEOETLGTIMPEESAKVCLDTVIMRFQELKQVERKIAQOAKIGIDIL  
 DTRVOQNOEKOKRHLDTVSSKIELNRLIDDOEOIOLKSTNELSKLOJST  
 NIOROOLEOETLESTEVOLSEIKDAKQVSPLETLLEKQOEKELINKKQJST  
 KIADKRLDKKKNHIGYKDIENYIDOKDQYKQKTELKVAQSECEKHE  
 KINEMRLIMRODIDTOKIOBRMLODNLTLEKREBELKEVEERKOLEKQOMLOM  
 KSRHOKLEENIDIKRNHNALROKQYEBEITHKEKLEDEQREKREKREKQVLM  
 RTTELVNKLDLITKTLDOAKIMKFSHKEEINKITDLKRSRQDIEYETRSBA  
 DENVSASDKRRNRYVMKIDPDLMDRGRCAGQGVASLIRLALATPCLNCI  
 IALDEPTNLDRENIESLAHALVEIIRKSROQNFOLVITHDEDFVELLGRSEYER  
 FYRIKKNIPOCSSEIVKCSVSLGFNVH"

BASE COUNT 1696 a 746 c 965 g 977 t  
 ORIGIN

alignment\_scores:                      Length: 1386  
                     Quality: 1662.00                      Gaps: 38  
                     Ratio: 1.878                              Percent Identity: 31.602  
                     Percent Similarity: 63.853

alignment\_block:  
 US-09-538-396-2 x AF057300 ..

Align seg 1/1 to: AF057300 from: 1 to: 4384

1 MetSerThrValAspLysMetLeuIleLysGlyLeuArgSerPhe..... 15  
 ||||| :|||:||||| ||| |||:|||||  
 322 ATGTCCCGAGATCGAAGATGAGATCTGGCGTGGAGATTGGAAAT 371  
 16 AsproAspAsnLysAsnValIleThrPhePheLysProLeuThrIleu 32  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||  
 372 AGAGACCAAGATAGCAAAATATCATCTTCTTCAGCCCTTACAAATTT 421  
 32 LevalGlyProAsnGlyValGlyLysThrThrIleIleGluCysLeuLys 48  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 422 TGGTGGACCCCAATGGCGGGAAGAGACCATCATTAATGATCTAAAT 471  
 49 LeuSerCysThrGlyLysLeuProProAsnSerArgSerGlyHisThrPh 65  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 472 TATATTGTACTGAGATTTCCCTCCTGGAACCAAA...GGAATATCAT 518  
 65 eValHisAspProLysValAlaGlyLysGluThrGlyLysGlyIleL 82  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 519 TGTACACATCCCAAGGTGCTCAAGAAACAGATGTGAGAGCCCAAGATTC 568  
 82 yslLeuArgPheLysThrAlaAlaGlyLysAspValAlaCysIleLeuArg 98  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 559 GCTGTCAATTTGTGATGTCAATGGAGAACTTATAGCTGTGCAAAAGATTC 618  
 99 PheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluL 115  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 619 ATGGCAT..... 625  
 115 ervaLleuGlnThrIleAsnProHisThrGlyLysValAlaCysLeuSer 131  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 647 TCTAAGTGTGGAATTTGACCGAGAAATGATCATGTTCTCTTGGGGTTTC 696  
 148 rLysAlaValLeuGluAsnValIlePheValHisGlnAspGluSerAsnT 165  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 697 CAAGCGTGTGCTAATAATATGTCATTTTCTGTCATCAGAGAAATTTCTAAT 746

165 rPProLeuGlnAspProSerThrLeuLysLysLysPheAspAspIlePhe 181  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 747 GGCCTTAAGTAGAAGAAAGGCTTGAAGCAAAATGTTGATGATGATTTT 796  
 182 SerAlaThrArgTyrThrLysAlaLeuGluValIleLysLysLeuHisL 198  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 797 TCAGCAACAAGATACATTAAAGCCCTTAAGAACCTTCGCGAGGTGCGCA 846  
 198 sAspGlnMetGlnGluLysLysThrPheArgLeuLysLeuGluAsnLeu 215  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 847 GACACAAAGCTCGAAGATGAAGAAATATCAATTAAGAACTTAATAATCTCGA 896  
 215 lnrThrValLysAspGlnAlaHisLysLeuArgGluAsnIleAlaAsp 231  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 897 ACCAATATTAAGAAAGAAAGCTGTGAGATTCTCGATGACATTAACAAGTAA 946  
 233 GlnGluLysSerAspAlaSerLysSerGlnMetGluGlnLeuLysGlyL 248  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 947 GAAGCCCAAGTTAACTTCAAGAGAAATTTGCAAAATCCTATGAGAAAGA 996  
 248 sIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAsp 265  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 997 ACTTGAATCCATTGAAGAAATCGCTAAAGAAATTAAGAACTTAATCTCTCA 1046  
 265 luleuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSer 281  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1047 AATAATTAAGAAATCTGACATGAATTAAGCTTGGATAGCGGAAGAAAG 1096  
 282 ThrLeuLeuThrGlnGlnHisGlyLysLeuAlaAlaLeuSerGlnGlu 298  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1097 CAATGAGAGAAAGATTAATGATGAGAAAGAAAGAAAGAAAGTTT 1146  
 298 nGluAspThrAspGluGluLeuMetGluThrIleThrLysPheGluGlu 315  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1147 TCAAGGAGCTAGTAGCACTTAATGACTATATGACATACCAAGAGAA 1196  
 315 rGleAlaLeuLeuGluGluThrLysIleSerLysLeuValArgAspMet 331  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1197 CAGTAAGGAGAAAGAAAGAAATGATGAGAAAGAAAGAAAGAAAGTAA 1246  
 332 AspGluAlaSerLysSerSerValLeuSerLysGlnAsnSerGluLeu 348  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1247 AACTAATAATTAAGAAATCTAGGCTCTCATATCGAAGAAATCGAAGCT 1296  
 348 rHisGluIleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThrMet 365  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1297 TGTGAAACGGCTCGCTACAGCTCAAGCAAGATGCCATCAAGAACTA 1346  
 365 ysnHisGluArgSerAspIleLysAsnIleCysThrLysHisAsnLeu 381  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1347 TCCGAGCTAGAGATTCATTAAATTCAGTCTTGGCAACAGCTAAGAAATG 1396  
 382 GlyProValProGluHisProPheThrAsnAspValAlaMetAsnLeu 398  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1397 GATGGCTTGAAGCTGAGCAATTCAGTGAAGAAACAGATTAATAATTTCA 1446  
 398 rAsnArgIleLysAlaArg.....LeuSerSerL 408  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1447 CAAGCTTGTGAGAGAGAGCAAGAAAGGAGACCAAAACTGCCAACCAAC 1496  
 408 euGlnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLeuAsp 424  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1497 TGATGAATGACTTTCGAGAAAGAGACTCTGAAACAAAGAAACAGATAG 1546  
 425 ValLeuThrLysHisThrLysLysIleAsnAlaArgTyrSerGluValAs 441  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1547 .....GAGATTAAGAGATTAAGAAAGAGAGCTGGG 1575  
 441 pGlyGlnIleGlnSerLysIleGluSerMetSerGlyLysLeuArgArg 458  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1576 AAGAAATTAATGATTAATCAAGAAATCCTAAGT.....A 1610  
 458 rGlyAspArgLysGluLysGluArgAspAlaAlaGluValGluLeuSerLys 474

```
1611 AGAAGCAGATGACGTGAAAAATCTGAGTATGATATACAGCAGCTTGGA 1660
      :::::::::::::: ::::: :::::
475 PheasnLeuSerArgIleaspGluArgGluArgHisMetGlnIleGluA 491
      :: ::::: ||| ::::: :::::
1661 GGATCTTCACAGAGATTTCTGAACCTGACAGAGAGCTC...ATTAAGC 1707
      :::::::::::::: ::::: :::::
491 LgluArgLysThrLeuAlaLeuGlyGluArgAspTyrAspSerIleLeu 508
      :::::::::::::: ::::: :::::
1708 TGAACGCTGAG...TTAAGCAAGCTGAGAAAAACAGCAATGTAG 1748
      :::::::::::::: ::::: :::::
508 ergLysArgThrGluValTyrSerLeuGlnLys... 520
      :: ::::: ||||| :::::
1749 AAACCTTAAAAATGAGTAAATAGTCTCCAAAATGAAAAAGCAGACTTA 1798
      :::::::::::::: ::::: :::::
521 .....IleLysValLeuLeuArgGluLysAspIleIleasnArgas 534
      ::::: ||| ::::: :::::
1799 GACAGACAGCTGCGTAAATGACAGAGATGGAGCAGTTAAACCATCA 1848
      :::::::::::::: ::::: :::::
534 nAlaaspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGlu 550
      ::::: ||| :::::
1849 TACACACACAGCTACCAAAATGAGATGTCACCAAGACAAAGCTGACA 1898
      ::::: ||| :::::
550 eSerLysAspLysLeuAsnGluIleValasnGlnHisLysAspLysIle 566
      ::::: ||| :::::
1899 AAGAT...GAACAAATCAGAAAAATTAATCTAGCGACAGTGAATTA 1945
      ::::: ||| :::::
567 LysLysValLeuArgGluArgAsnProPheGluLysAspMetLysGlu 583
      ::::: ||| :::::
1946 ACCTCACTGTG...GGATATTTCCACAAAAAACAGCTGAAAGAC.. 1990
      ::::: ||| :::::
583 uIleasnGlnAlaPheTrpProValAspLysGluTyrAsnGluLeuArg 600
      ::::: ||| :::::
1991 .....TGC.....CTACATA 2000
      :::::::::::::: :::::
600 eLysSerGlnGluAlaGluGlnLeuLysPheThrGlnSerLysVal 616
      ::::: ||| :::::
2001 GTAAATCAAAAGAAATTAATCAG..... 2023
      :::::::::::::: :::::
617 ThrAspAlaArgGluGlnLeuThrLysLeuArgArgAspMet..... 630
      ::::: ||| :::::
2024 .....ACCAGGACAGACTGTCCAAAATGAAACAAAGACTGATCATC 2067
      :::::::::::::: :::::
631 .AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuG 646
      ::::: ||| :::::
2068 TGACACAGAAATTAATCATATTAATGAACCTAAAAAGAAAGAGAGC 2117
      ::::: ||| :::::
646 InIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAsp 660
      ::::: ||| :::::
2118 AGTGTCCAGTTACGAAAGACAAGCTGTTGATGTTGTGTGTAAGCAGAT 2167
      ::::: ||| :::::
661 AlameLysArgAspGluGlnLysArgLeuGlnLysPheAlaAsnGlu 677
      ::::: ||| :::::
2168 TTGTAAGTGTATTGACAGGCTTAAGAGGAATGAAAAATCATCAAA 2217
      ::::: ||| :::::
677 yMetArgGluMetLeuAlaPro.....PheG 686
      ::::: ||| :::::
2218 ACACGAGCCCTGTTGGCTGAGCCACAGCAGTTACTCCAGTTCATTA 2267
      ::::: ||| :::::
686 LuHisLeuAlaArgLysAsnHis...ValLysProCysGlyAlaAla 701
      ::::: ||| :::::
2268 CTCAGCTAACAGACAAACAGCATGTTGCCCTGTTGTCAGAGAGTT 2317
      ::::: ||| :::::
702 Phe...ThrProAspGluGluAspGluPheValLysLysGlnArgMetGlu 717
      ::::: ||| :::::
2318 TTTCAGACAGAGGCTGAGTTACAAAGAGTCACTGATTTGCGAGTCTAA 2367
      ::::: ||| :::::
717 nAsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetClnSerSerA 734
      ::::: ||| :::::
2368 ACTCGCACTTGCTCCAGATTAACCAAGTCAACCAAGTCAAGCTTAATA 2417
      ::::: ||| :::::
734 snAlaGluAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleLysAsp 750
      ::::: ||| :::::

2418 AAAAGAAAAAGCGCGTATGTAATGCTGGGACTTGCCCATGAGCAA 2467
      :::::::::::::: :::::
751 AlaTyrValLysLeuValGluGluThrIlePro...LeuAlaGluLys.. 765
      ::::: ||| :::::
2468 AGCATTAATGATTTGAAGAGAAAGAAATATCCAGAAATTAAGAAACACT 2517
      ::::: ||| :::::
766 ....AsnLeuasnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaP 781
      ::::: ||| :::::
2518 GCAGATGTCAATGACACATACAGCCCTTAAGAACAGACATTAAGAAC 2567
      ::::: ||| :::::
781 heAspAspLeuLeuGly..... 786
      ::::: ||| :::::
2568 AAGAAACACTCTTGTTGTAACAATAATGCCGTAAGAAAGTCCAAAGTA 2617
      ::::: ||| :::::
787 .....ValLeuAlaHisValGlnMetAspArgAspAl 797
      ::::: ||| :::::
2618 TGCCCTGACAGATGTTACAAATTATGAGAGGTTCCAGATGCAACTTAAGA 2667
      ::::: ||| :::::
797 aValGlu.....AlaLeuLeuGlnProThrAspT 807
      ::::: ||| :::::
2668 TGTTGAAAGAAAAATTCACACAAAGCAGCTTAAGCTACAAAGATACAC. 2716
      ::::: ||| :::::
807 hIleAspArgHisValHisGluIleGlnGlnLeuValLysGluValGlu 823
      ::::: ||| :::::
2717 ..TAGATCGAAGCTGCCAACAGTCAACGAG.....GAGAAACAA 2755
      ::::: ||| :::::
824 AspLeuGluTyrAlaLeuAspSerSerGlyArgGlyValLysSerLeuG 840
      ::::: ||| :::::
2756 GAGAAACAGCAGCAAGTTAGACACA.....GTTCTTACTAAGAT 2793
      ::::: ||| :::::
840 uGluIleGlnLeuGlnLeuAsnPheLeuGlnArgThrArgAspThrLeu 857
      ::::: ||| :::::
2794 TGAATGAATGCTAGCTT..... 2812
      ::::: ||| :::::
857 LeValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGluAsp 873
      ::::: ||| :::::
2813 .....ATACAGAGCAGCAGAAAGATCAACATCA 2845
      ::::: ||| :::::
874 MetSerSerAlaGlnValArgTrpHisAsnAlaArgGluGluLysValL 890
      ::::: ||| :::::
2846 AAAAGTCAACA.....AATGAGCTAAAACTGGAACACTCA 2883
      ::::: ||| :::::
890 sAlaSerSerIleLeuGlnArgPheGlnLysSerGluGlnLeuValL 907
      ::::: ||| :::::
2884 GATATCCACTAATTTGCAACGCTGCAACACTGAGAGAGAGCTGTGG 2933
      ::::: ||| :::::
907 euleuAlaGluGluLysGlnLeuIleValGluLysLysLeuGlu 923
      ::::: ||| :::::
2934 AATTAATCCAGTGAAGTCACTGTTGTACAGAGATTAAGATGCTAA 2983
      ::::: ||| :::::
924 GluSerLeuAspProLeu.....SerLysGluLys 933
      ::::: ||| :::::
2984 GAGCAGGTTAAGCCCTTGAAACAAACATTGCAAAAGTTCCAGCAAGAAA 3033
      ::::: ||| :::::
933 sGluSerLeuLeuGlnGluTyrAsnAlaLeuLysGlnLysLeuAspGlu 950
      ::::: ||| :::::
3034 AGAAGAAATTAATCAACAAAAAATACAGCACAAATACACAGAGGTA 3083
      ::::: ||| :::::
950 LuTyrHisGlnLeuAlaGluArgLysArgGluPheGlnGlnLeuAsp 966
      ::::: ||| :::::
3084 AACGTGAATGATTAAGAGAGGTTAAATATTCATGCGTATATG... 310
      ::::: ||| :::::
967 AlaLeuGlyArgLeuAsnMetLysIleLysGluTyrLeuAspSerLysL 983
      ::::: ||| :::::
3131 .....AAAGACATTGACAAATATTAATTCAGATGAGAA 3162
      ::::: ||| :::::
983 sAsnGlu.....LysLeuLysGluLeuGlnGluArgHisValLeuCysH 998
      ::::: ||| :::::
3163 AGACGACTATTAAGCAAAAAAGAACTGAACCTTAATTAATTAATA... 3208
      ::::: ||| :::::
998 IsSerGlnLeuGlnSerCysMetalLysGlnGlnArgIleSerAlaGlu 1014
      ::::: ||| :::::
3209 ..GCTCAACTAGTGAATGCGAAGAAACAAAGAAAGATTAATGAAGAT 3256
      ::::: ||| :::::
```

1015 LeuasnlySerlySgluLeuclnglnclynglInleuLeuArgas 1031  
 3257 ATGACACTCATGACAGAGATATTGATACAGAGATACAGAAAGGTG 3306  
 1031 nllspaspasnleuLytyrArglyThrlyAlaSpValglnglInL 1048  
 3307 GCTACAAATACCTTACTTACAGAAAAAGAAATGAGCACTAAAGAG 3356  
 1048 eutHrArgAspIleGluSerLeuGluArgLeuSerIleGlySer 1064  
 3357 TTGA.....GAGAAAGAAACAACATTGAGAAATGGGTCA 3397  
 1065 LeuSerAlaIleGluAlaSpLeuLyArgHisSerGlnGluLys... 1079  
 3398 ATCGAGTTTGCAATGAAAAAGTACATCAGAGTTGGAGAGAACAT 3447  
 1080 GluArgLeuasnSerGluPheAsnArgTyrPglnglInlyThrLeuSerValT 1096  
 3448 AGACATATATAAGAAATCATATTTGGCATTAGGGCCAGAAAGGTT 3497  
 1096 YrGlnSerAsnIleSerLyHisLySglnglInleuLysLeuSerGlnTyr 1112  
 3498 ATGAGAGAAATATTTCATTTTAAAGAAAGACTCGAAGACACATTT 3547  
 1113 LysAspIleGluLysArgTyrThrAsnGlnPheLeuGlnLeuLysThrTh 1129  
 3548 CGGATGCTGAGGAAAGATAGAAATGATGATTTATGAGCAAC 3597  
 1129 RGlueMetAlaSnlySaspLeuAspArgTyrTyrThrAlaLeuAspLysA 1146  
 3598 AGAATCTTGAAACAAGATCTCGATTTATTTAAGACTCTTGACCAAG 3647  
 1146 IalMetLarPheHisSerMetLysMetGluGluIleAsnLysIleIle 1162  
 3648 CAATATGAAATTCACACTATGAAATGAGAAATCATATAATTA 3697  
 1163 LysGluLeuTyrPglnglInThrTyrArgGlyGlnAspIleAspTyrIleSe 1179  
 3698 CGTGCAGCTGGCGAAGTACCTATCTGACAAATATTGATACATAGA 3747  
 1179 rIleAsnSerAsp.....SerGlnGlyAlaGlyThrArgS 1191  
 3748 AATAGGCTGATGCCGATGAAATGATCATCTGATTAAGGCGGA 3797  
 1191 eTyrSerTyrArgValMetGlnThrGlyAspAlaGluLeuGlnMet 1207  
 3798 ATTATTAACCGAGTGTGATGCTGAAGGAGACACACCTTGGATAG 3847  
 1208 ArgGlyArgCysSerAlaGlyGlnLysValLeuAlaSerLeuIleLea 1224  
 3848 CGAGACAGATGAGTGTGACAAAAGGATATAGCTCATCATCATTCG 3897  
 1224 GLeuAlaLeuAlaGluThrPheCysLeuAsnGlyIleLeuAlaLeuA 1241  
 3898 CCTGGCCCTGGTGAAGCTTCTGCCCTCAACGTGGCACTGATGGCTGG 3947  
 1241 sPglProThrThrAsnLeuAspGlyProAsnAlaGluSerLeuAlaA 1257  
 3948 ATGAGCCAAACAATCTTGACCGAAGAAACATGAACTCTTGACACT 3997  
 1258 AlaLeuLeuArgIleMetGluAlaArgLysGlyGlnGluAsnPheGlnLe 1274  
 3998 GCTCGGTTGAGATATAAAGTCGCTCAGACGAGCGTAACCTCCAGCT 4047  
 1274 uIleValIleThrHisAspGluArgPheAlaHisLeuIleGlyGlnArg 1291  
 4048 TCTGCTATCACTCATGAGGATTTCTGAGCTTTTGAAGACGTTCTG 4097  
 1291 InLeuAlaGluLysTyrTyrArgValSerLysAspGluAsnGlnHisSer 1307  
 4098 AATATGTGAGAAATTTACAGGATTAAGAAACATCGATCAGTGTCTCA 4147

1308 IleIle 1309  
 4148 GAGATT 4153

seq\_name: .9b\_pl3:SCRAD50

seq\_documentation\_block: 4775 bp DNA

LOCUS SCRAD50 4775 bp DNA PLN 12-SEP-1993.

DEFINITION Yeast RAD50 gene for 153 kD protein.

ACCESSION X14814

VERSION X14814.1 GI:4272

KEYWORDS DNA repair; DNA-binding protein; meiotic recombination; RAD50 gene.

SOURCE Saccharomyces cerevisiae.

ORGANISM Saccharomyces cerevisiae.

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

AUTHORS 1 (bases 1 to 4775)

TITLE Direct Submission

JOURNAL Submitted (21-MAR-1989) Alan E., Harvard University, 7 Divinity

REFERENCE Avenue, Cambridge MA 02138, U S A

AUTHORS 2 (bases 1 to 4775)

TITLE Alan E., Subbiah, S. and Kleckner, N.

REFERENCE The yeast RAD50 gene encodes a predicted 153kd protein containing a

purine nucleotide binding domain and two large heptad repeat

regions

Genetics 112, 47-57 (1989)

3 (bases 1 to 4775)

REFERENCE Raymond, N.E. and Kleckner, N.

AUTHORS Expression of the Saccharomyces cerevisiae RAD50 gene during

meiosis: steady-state transcript levels rise and fall while

steady-state protein levels remain constant

Mol. Gen. Genet. 238 (3), 390-400 (1993)

93261422

DATA kindly reviewed (25-SEP-1989) by Alan E.

Location/Qualifiers

1..4775

/organism="Saccharomyces cerevisiae"

/strain="REB21"

/db\_xref="taxon:4932"

/chromosome="chromosome 14"

/clone="pSG205"

532..541

misc\_feature

CDS

/note="region of transcription start"

558..4496

/note="153 kD protein (AA 1 - 1312)."

/codon\_start=1

/protein\_id="CAA32919.1"

/db\_xref="GI:4273"

/db\_xref="SWISS-PROT:P12753"

/translation="MSAIVKLSIGISFDSNDRETFEGRPYLIVGNGSGRTII

ECLKATATGDLPPRSKGVFIHPDKTIGEDIRAOVLAFTSANGNMYRNIOILM

KTTTPEKTELEGOLVAINNGSDRSITSTRELDAOVPLVGLVPAVILEVYICHD

SLNPLESPNWKRFDFIRPAKFTKLDLTKSTKMSDIDILKQSVHNLKDR

SKAKNLINHLQRTKIDYNEVESEISQLNEITEKSDKSNQDQKILSYENLKN

TKLISDVQVRKLSNIDILSKPDLQNLNLANFSYMDNNQDLBLEDDISLKRO

SLLOSLSNLSIRGSELAGETEKERNILSKLEAFQFGLNKEINSDSAQVNH

EMSOFKAFITSGDLDTIDDFAKDILKETNLSDIKISIVDSQNLNLEKRNKLIH

SEBAETLKSPKSLSTQDSANHELENKTKREKLOSSESNTIPIKLNKKEENNM

ILENOLKFDRLIKTNQADVLAKGLIKSTYTKLDELQKTEKLNDSRPYFP

LTOEFQRADEMPQKLFIMQKNIALNNKMEHLDRTYNALYNTLEKDLQDNK

SKREVIOLESENPECDTIDYNEVESEISQLNEITEKSDKSNQDQKILSYENLKN

RKSCCYLSKRKFENESKSLQELKRTDANETKSTYKTLNENKMQTLLEFRKLAL

ITLNSINEKIDNSQKLEKKEETKSKSLDELEVOSTKDEKELALSEIPLLEK

PTYLEKLEKLENSKRTISELSYNSSEGIQVPELPOORKNDSRLKELKTSID

LOMKEDKVRKNSKMLIKELKELTYSEISSLQKQNIQDSIRSRKNINDSDSVK

ELERKISLNKRNKQAGSVLDKYNEDIDVRRNKQVADINILDRPOTTYEVDF

ENKGFDELQTTIKELIELEKQMLKEDQDLKSNVEERKRLADSNNEKRNKQML

LIKESOLOHIESEISRLDVONAEADKQOESLRTAFREKLSSENAKGLKQML

ONQDLSLHLDRTDYKDIERNYKHEWELOTRFVMDIDVYSKALDSAIKYGHLK

ODIRITDELKRTYSGTDITIKIRSDVYSYKGSYRYRYVMKODEYEDMRKQ

SAGKVLASITIRLALSETFGANGVIALDEPTTNDEENIESLANSNNTIMRRHQ

```
misc_feature      642..707      KNFOLIVITHEDEKELGHNNAAFIDHEFKVRDRKQSOIEWVDINRTVY
misc_feature      1089..1820    /note="nucleotide-binding domain"
misc_feature      2784..3542    /note="heptad repeat"
misc_feature      /note="heptad repeat"
BASE COUNT      1827 a      768 c      895 g      1285 t
ORIGIN
alignment_scores:
  Quality: 1477.50      Length: 1404
  Ratio: 1.666          Gaps: 41
  Percent Similarity: 63.177      Percent Identity: 29.701
alignment_block:
  US-09-538-396-2 x SCRAD50
Align seg 1/1 to: SCRAD50 from: 1 to: 4775
1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPheAsp 17
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
558 ATGAGCGCTATCTATAATATCTATCATGAGGCAATACGGCTTTGACTC 607
17 caapaenlysaasnvallethrphelhyserproleuthrleuvala 34
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
608 CAATGATAGGGAACCTATGATTTGGCAACCTCTGACTTAAATGTCG 657
34 LProAsnGlyAlaGlyLysThrThrIleIleGlyLysLeuLysLeuSer 50
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
658 GCATGATGATGTCAGTAAACAGCATATCATCATGATTTAAAGTACGCT 707
51 CysThrGlyLysLeuProAsnSerArgSerGlyHisThrPheValHis 67
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
708 ACCACAGGTATCTACCGCCCAACAGCAAGAGGAGA...GTATTCATTC 754
67 saaprollysaalalaaglyluthrghluthrlysglylnlleuleua 84
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
755 TGACCCGCAATATCTGCTGAGAAAGCATTAAGCCTCAGCTCAACCTCG 804
84 rphelhystrhlaalaglylysaaspyalaCysIleArgSerPheGln 100
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
805 CGTTTGAGATGCGCAATGACATATGATGTGCACCAAGAAATATATCAG 854
101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerValle 117
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
855 TTCTATATGAAAAGACTACTACTACTATTAGACTTTAGAGCCAGCTT 904
117 uGlnThrIleAsnProHisThrGlyGlyLysValCysLeuSerThrArg 134
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
905 GGTGCGCTATTAAT...AATAGTGTGACCGCAGTACTTGTCCACCCGGT 951
134 ysAlaAspMetAspArgGluIleProAlaLeuMetGlyValSerLysAla 150
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
952 CTCGTGATATAGACGACAGTACCCCTGATCTGGAGAGTTCCAAAAGCG 1001
151 ValLeuGlnAsnValIlePheValHisGlnAspGluSerAsnTrpProle 167
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1002 ATCTTGGAATACGTGATATCTGTCCAAAGAAAGACAGCTGTGGCCACT 1051
167 uGlnAspProSerThrLeuLysLysPheAspPheIlePheSerAlat 184
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1052 GAGTGAACCATCCAAATCTTAAAGAAAGTTGATGAATTTTCCAAAGCGA 1101
184 hrArgTrpThrLysAlaLeuGlnValIleLysLysLeuHisLysAspGln 200
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1102 TGAATTTAAAGAGCTGACATTTAAATCCATTAAGAAAGAACATG 1151
201 MetGlnGluIleLysThrPheArgLeuLysLeuGlnAsnLeuGlnThrVa 217
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1152 TCCGTGATATCAAGTTATTAAGCAATAGTAGAACACTTAAGTTG... 1199
```

```
217 LysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnLul 234
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1200 .....GATTAAGACA 1209
234 ysSerAspAlaSerLysSerGlnMetGlnLeuLysGluLysIleCys 250
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1210 GATCAAGAACCAATGAAAGCTCAATATCCCAATGCAAACTAGATGAT 1259
251 GlyThrGlnArgGluIleLeuGlnMetGlnThrSerLeuAspGluLeu 267
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1260 CAGTATATGAAAGAGAGTGTGAATTTGCAATCTCAATGAATGAATC... 1307
267 ArgLeuGlnGlyGlnIleAspIleLysAlaThrGlnArgSerThrLeu 284
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1308 .....ACTGAGAAATCCGATTAAC 1326
284 eu.....ThrGlnGlnHisGluLysLeuAlaAlaLeuSerGluGln 297
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1327 TGTTTAATCTAATCAAGATTTCCAAAATACTATCCAAAGTAGAGAAC 1376
298 AsnGluAspThrAspGluGluLeuMetGluTrpGlnThrLysPheGln 314
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1377 TTGAAAATCAAAACCTCAATTTAGTATCAAGTGAAGAGATTGTCAA 1426
314 uArgIleAlaLeuLeuGln.....T 321
1427 CTCATTTGATATTTCTCGATTTATCGAAGCCAGATTTCCAAACTTATG 1476
321 hrLysIleSerLysLeuValArgAsp.....MetAspAsp 332
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1477 CAATCTCTCTAAGTCTGATGATGATGAATATATCAGTTGAGAGATTG 1526
333 GluAlaSerTyrosSerValLeuSerLysGlnAsnSer.....GluLe 347
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1527 GAAACCGATTTTCAAGTTGAAAGACCGTCATCATCCTTGACAGACCT 1576
347 uThrHisGlnIleGlyLysLeuGlnAlaGlnAlaAspAlaHisLeuThr 364
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1577 TTCTAATTCCTCATCGCGCGCAGCGCTGAGTTAGAGCT..... 1616
364 euLysHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsn 380
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1617 .....GCTAAGAGACATACGAAAGAAAGCAAGAAATCAGCTTTCATG 1658
381 LeuGlyProValProGlnHisProPheThrAsnAspValAlaMetAsnLe 397
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1659 CTAAGAAACACCTTCCAAACATAGTTC.....CAAGGCT 1693
397 uThrAsnArgIleLysAlaArgLeuSerSerLeuGlnAsnAspLeu 414
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1694 TTCAAATATTGAAACAGTATATGCTCAAGTAAATCAGCAATATGTCAC 1743
414 sp...LysLysLysSerAsnGluAspGlnLeuAspValLeuTrpLysHis 429
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1744 AGTCAAGCTTCATCATCTCAGAGATTACTGACACTATAGATCAATTT 1793
430 TyrLeuLysIleAsnAlaArgTyrosGluValAspGlyGlnIleGlnse 446
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1794 GCAAGAGCTTTCACATTAACAAATTTATCCCACTTGATTAATATC 1843
446 LysIleGluSerMetSerGlyLeuLeuArgArgArgLysAspLysGlnLul 463
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1844 AATTAACCGTGTGATTCACAGATTTGCAATTAATTAAGAGACGATCTA 1893
463 ys.....GluArgAspAlaAlaGluValGluLeuSerLysPhe 475
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1894 AATTAATTCATGACTCGAAGAGCTTGCTGAGAAATGAAATCATTTAA 1943
476 AsnLeuSerArgIleAspGluArgGluArgHisMetGln...IleGluVa 491
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1944 AGTTGCTCTCAAGATACCTCAATCATCAATGAACTGAAAACTTAAGAC 1993
491 LgluArgLysThrLeuAlaLeuGlyGluArgAsp.....T 503
```



1994 ATACAAAGAAAGCTTCATCTTGGAAAGTGAAATATTATTCAGAAAT 2043  
503 yrasperilleeserleuglnlyargthgluvaltyrserleugln 519  
2044 TGACACGAAATAGAGAGAAATATGATGATCTTACAGAAAT 2093  
520 lyslileysvalleuarglulysasplleileasnargasnlaas 536  
2094 CAATTTGAGAAATTCAGATATGAAATGAAACGAAATCAACACGGG 2143  
536 pgluargvalysleuglyleuylsyrasplaleugluserleu 553  
2144 CTTATACGCAACTAGACTATTAAGAAGTCCATTAATACAAATTAG 2193  
553 splysleuasnglulilevalasn.....gluhslyaspllyle 567  
2194 AGAGATTCAAAAGATTACGAAATATACAGATGATCTACAAATAG 2243  
568 lysvalleuargglyargasnpropheglulys..aspmetyls 583  
2244 CAAGTATTTCCCTCACCGAGATTTCAAGAGCTGATTTAGAAATGGA 2293  
583 uileasngluinalaphetprowalaspysgluylasngluleuarg 600  
2294 TTTTCGAAATATTATTCATATGCAATATGCAAGATATAGCATCACA 2343  
600 eyleserleuglnlualaglulnglululeuylserthrglnser 616  
2344 AAAAGTGCACGAAATGATAG.....AGATAT 2372  
617 thrasplalargglulngluleuthlyleuargargaspmetasp 633  
2373 ACAAAAGCTTTGTACATTTGAAACACCATTTGAAAGATTTACAGAT 2422  
633 sargargpheleuasperlyleuglnserle.....leugln 647  
2423 TCAGAGTGCAGAAAGAAAGTGAATACACTACACGAAACTTACAG 2472  
647 leserlaasnvalaspmetpheprolyvalleuglnasplamet 663  
2473 AGGATGTACTATGATGATATATGATGATCTCTGAGAA.....ACG 2516  
664 lysargaspglulnglulysargleuglnasplaleuarglymet 680  
2517 GAGCTCTCTTATAGACTGCTGAGAAACTGAAA.....ATGCATCA 2560  
680 uileleuallapropheglu.....hileuallarglyasn 693  
2561 AACTACTTAGAATTTAATAGAAAGCATTAGAGATCGCTGAGACGT 2610  
693 lavalcyprocyscysgluargalaphetprowaspgluarglu 709  
2611 GGTGCTGATTTGCTGTCTAGAAATTT.....GAAATGATCG 2651  
710 phevallyleuys.....gluargmetgluasnserlethralglu 724  
2652 TTTTAAAGTAAAGTGTTCACAGAACTAAACAAACAGACCCCAATT 2701  
724 gserlysalaleuallametgluseraserasnaglualaleu 741  
2702 TGAATAAACTTTGAAAGACACTGTCGAAACGAAAG.....GAAT 2742  
741 luleuaspysleuargthlylethrasplalatyvallyleu..... 755  
2743 ATCTTCATAGTTTAGGCTACTGAGAAACATTTATACACTAAATTC 2792  
756 valgluargthlyleuargleuarglyasnleuasnghileuall 772  
2793 ATAAAGTAAAGTGTATATTCACAGAAAGTCCCTGAAACCACTAAG 2842  
772 aspgluserglulnglulnglualaphesapleuarglyvalleu 789

2843 GGAACGAAAACTTCCAAATCAAACTAGATGAATGGAAGTATTC 2892  
789 lahivalgluinetasparthargaspalavalglualaleu 805  
2893 CAATAATGAAAGATGAGAAAGAGTGTGAGCTGATTAATCTCTTGG 2942  
806 asptthrlasparthvalhlsghlueuglnleuvallyleu 822  
2943 .....ATTGAGAGTGTACTTACTTGAAGAAAGT 2974  
822 lgluaspleuglulthralaleuaspersergylargglyvally 839  
2975 TAGGATCTTGA.....AACGTT 2994  
839 euuglulileuglnleuasnphleuglnargthargapthr 855  
2995 CTAAACAGATCTGTGAGAGCTATTCATTTATATACGATGAGTGT 3044  
856 leuilevalgluvalasparleuargaspghlueuargmetleu 872  
3045 ..ATTCAACCTGTGTATGATTAATTAAGGACCAACAAAGAGTGAAT 3091  
872 uaspmet.....sereralnglinalargtrp 882  
3092 TTCTTACGTAATGAGGAAACATTTACGATTTGCAATGGA..... 3137  
882 lsaasnlaarggluarglyvallyalserleleuglnargp 898  
3138 .....AAAGATGAAAGGTGAGCAAAATTCAGAAATGATTAATCT 3179  
899 glulysserglulngluleuvalleuallaglulnglulngl 915  
3180 ATTAAGAAAGAA.....TTTACAGTTTGTGAATTTGAATCAATC 3220  
915 uilevalglulysleuuglnuser..... 925  
3221 TTTTACACGAAACAAATTTATGATCTTATTAAGTCAAAAGGAA 3270  
926 .....leuasperleuaserlysglulysgluserleu 938  
3271 ACATCAACGATATTCGATTTAGAGTAAAGGATTAACACCTAT... 3317  
939 gluthrasnlaaleuysglulysleuaspglulthylhlsghleu 955  
3318 .....ATTTCATTAATAAAACAAAGAGTAAAGCCCAAGTCTTGA 3361  
955 agluarglys.....arglupehnglulngl 965  
3362 CAAGTAAATGAACGATATTCAGATACCAACAAACAAACCG 3411  
965 euaspalaleuarglyleu..... 971  
3412 TTGCAGATATTAATCGTTAATAGATGATTTACAGCAATTTAAGAA 3461  
972 .....asmetlyslileysglytyrleuasperlyleuasn 986  
3462 GTGGTGTATTTGAACTTAAAGGCTTC.....GAAGAAATGGAACAA 3505  
986 sleuylsglueuglnlyarghlsvalleucysghlsseuglnleu 1003  
3506 AATAAAGAACTGAA.....TTGAACAAAGCTCAAAATCTG 3543  
1003 efcysmetalauglulnglularglileseralaglueuasnly 1019  
3544 AACTAAAGGAGAGCTGATTCAGCAAAATGAGTTAAGCAAGAAAG 3593  
1020 gluleuuglnlglulnglulnglulargthargasnle 1036  
3594 AGAAACTGCTGATCAATTAATGAGGAAACAACTTGAAGGAGACT 3643  
1036 ulyslyarglyserthlysalasplalaglulnglulnglularg 1053  
3644 AGAATTAATGATTTGAATCACAACCTTCAGCATTTGAATCGAATA 3693

[illegible][illegible]

17 oAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuIleValG 34  
1145 CAATGATAGGAACTATTGAAATTGGCAAGCTCTGACTTTAAATGTCG 1194  
34 LProAsnGlyValIleGlyLysThrThrIleIleGlyLysLeuLysSer 50  
1195 GCATGATAGTCTCAGSTAAAGACATCATCGAATGTTAAAGTACGCT 1244  
51 CysThrGlyLysLeuProAsnSerArgSerGlyHisThrPheValH 67  
1245 ACCACAGGATCTACCCGCCCAACCAAGCAGGAGA...GTATTCATTTCA 1291  
67 sAspProLysValAlaGlyLysLeuThrLysGlyLysIleLeuLeu 84  
1292 TGACCCGGAAGATTAAGTGTGAAAAGACATTAGAGCTCAGCTAAACTGG 1341  
84 rPheLysThrAlaIleGlyLysAspValAlaCysIleArgSerPheGln 100  
1342 CGTTTACGAGTCCCAATGAGCTCAATATCATCTCCACGAAATATTCAG 1391  
101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerVal 117  
1392 TTGCTAATGAAAGAAAGACTACTACTATTAAAGACTTGAAGCAGCT 1441  
117 uGlnThrIleAsnProHisThrGlyLysValCysLeuSerTyrArg 134  
1442 GGTGCTGATTAAT...AATAGGTGTGACCGCAGTACTTGTGTCACCCGG 1488  
134 yAlaAspMetAspArgGluIleProAlaLeuMetGlyValSerLysAla 150  
1489 CTCTGGAATTAAGCCACAGACAGTACCCCTGATCTGGAGTCCAAAGCG 1538  
151 ValLeuGluAsnValIlePheValHisGlnAspGluSerAsnTyrPro 167  
1539 ACTTGGAGTAGCTGATATTCTGTCTACCAAGAACAGCCTGGCCACT 1588  
167 uGlnAspProSerThrLeuLysLysPheAspIlePheSerAla 184  
1589 GAGTGAAACATCCAACTTAAAGAAAGTTGATGAATTTTCCAAAGCGA 1638  
184 hArgTyrThrLysAlaLeuGluValIleLysLysLeuHisLysAspGln 200  
1639 TGAATATTCAAAAGCTCTAGACAATTTAAATCCATAAGAAAGACATG 1688  
201 MetGlnGluIleLysThrPheArgLeuLysLeuGluAsnLeuThr 217  
1689 TCCGTGATATCAAGTTATTAAAGCAATCAGTAGAACACTTAAAGTTG.. 1736  
217 lLysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnGlu 234  
1737 .....GATTAAGACA 1746  
234 ySerAspLysSerLysSerGlnMetGluLeuLysGluLysIleCys 250  
1747 GATCGAAACAAATGAAGCTCAATATCCACCACAAATGCAAACTAAGATGAT 1796  
251 GlyThrGluArgGluLeuLeuGlnMetGluThrSerLeuAspGluLeu 267  
1797 CAGTAAATGAAGAGACTGTCTGAATTCATTCATTCATGAATGAATC.. 1844  
267 gArgLeuGlnGlyLysIleAspIleLysAlaThrGluArgSerThr 284  
1845 .....ACTGAGAAATCCGATTAAC 1863  
284 eu.....ThrGlnGlnHisGluLysLeuAlaIleLeuSerGluGlu 297  
1864 TGTTTAATCTAATCAAGATTTCCAAATAATCTATCCAAAGTAGAGAAC 1913  
298 AsnGluAspThrAspGluLysLeuMetGluTyrGlnThrLysPheGlu 314  
1914 TTGAAATAATACAAACTCTCAATTAAGTATGATCAAGTAAAGATTTGCAAA 1963

314 uArgIleAlaLeuLeuGlu.....T 321  
1964 CTCAATTTGATATCTGATTTATCGAAGCCAGATTTGCCAAACTATTG 2013  
321 hLysIleSerLysLeuValArgAsp.....MetAspAsp 332  
2014 CAATCTCTCTAAAGCTTGATGATAGAAATGATCAAGTATGAGATTTG 2063  
333 GluAlaSerTyrSerValLeuSerLysGlnAsnSer.....GluLe 347  
2064 GAAACCGATTTTCAAGTTTGAAAGACCGCATCTCATCTGCAAGACCT 2113  
347 uThrHisGluIleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThr 364  
2114 TTCTAATCTCCCTCATCAGCGCGCAGGTGATGAGAGCT..... 2153  
364 eLysHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsn 380  
2154 .....GGTAAGAGACATACGAAAGAAACAGAAATCACCTTTCATCG 2195  
381 LeuGlyProValProGluHisProPheThrAsnAspValAlaMetAsn 397  
2196 CTAAAGAAAGCTTCCACATTAAGTTC.....CAAGGCT 2230  
397 uThrAsnArgIleLysAlaArgLeuSerSerLeuGluAsnAspLeuVal 414  
2231 TTCAAATATTGAAACAGTATATGGCTCAAGTAAATACAGAAATGTCAC 2280  
414 sP...LysLysLysSerAsnGluAspGlnLeuAspValLeuTyrHis 429  
2281 AGTTCAAGCTTTCATCTCAGATTTAAGTCAAGTAAATACAGAAATGTCAC 2330  
430 TyrLeuLysIleAsnAlaArgTyrSerGluValAspGlyGlnIleGln 446  
2331 GCAAGAGATTTCACTAAGAAAGAAACAAATTTTCCGACTGTTAAATAC 2380  
446 rLysIleGluSerMetSerGlyLeuArgArgArgLysAspGlu 463  
2381 AATTAACCGTGATCACAGATTTGGAAATTAATTAAGAGACAGATCTA 2430  
463 yS.....GluArgAspAlaAlaGluValGluLeuSerLysPhe 475  
2431 AATTAAATCCATGACTCGAAGACCTTCTGGAATAATTAAGAAATCTTAA 2480  
476 AsnLeuSerArgIleAspGluArgGluArgHisMetGln...IleGluVal 491  
2481 AGTTTGCTCTCAAGTATCCCAATCAATCAAGTATGAAATCTTAAATAC 2530  
491 lGluArgLysThrLeuAlaLeuGlyGluArgAsp.....T 503  
2531 ATACAAAGAAAGCTCGAATCTTGGGAAGCTGGAATATTATCCAAAT 2580  
503 yAspSerIleIleSerGlnLysArgThrGluValTyrSerLeuGluGln 519  
2581 TGAACCGAAATATGAAGAAATAATATGATGATTAATCTTAAGAT 2630  
520 LysIleLysValLeuLeuArgGluLysAspIleIleAsnArgAsnAlaAs 536  
2631 CAATTTGAAGAAATTTCAAGATCCCAATATGAAGAAACGATCAACAAACG 2680  
536 pGluArgValLysLeuGlyLeuLysLysAspAlaLeuGluSerSerLys 553  
2681 CTATACGCCCAACTAGACTAATTAAGAAAGTCAATTAATCAATTAAG 2730  
553 sPlyLysLeuAsnGluIleValAsn.....GluHisLysAspLysIleLys 567  
2731 ACAGATTACAAAGATTTACGAAATAATTAACAGAAATGTTTAAATAAG 2780  
568 LysValLeuArgGlyArgAsnProPheGluLys...AspMetLysGln 583  
2781 CAAGTATTTCCCTCAACCCAGAAATTTCAAGAGAGATTTAAGAAATGCA 2830  
583 uIleAsnGlnAlaPheTyrProValAspLysGluTyrAsnGluLeuArg 600

```

2831 TTTTCAGAAATATATTCATCAATATGCAAAAGAAATATAGGATCAACACAA 2880
600 eTlysSerGlnGluIleGlnGlnGlnLeuLysPheThrGlnSerLysVal 616
2881 AAAAGATGCACGAATTTGGATGAA.....AGATAT 2909
617 ThrAspAlaArgGlnGlnLeuThrLysLeuArgArgAspMetLysPheVal 633
2910 ACAAAAGCTTTGTGACAAATTTGAAACACATGAAAGATTTACAGATGAA 2959
633 sArgArgPheLeuAspSerLysLeuGlnSerLys.....LeuGlnI 647
2960 TCAGAGAGTCGAGAAAGAAAGATGATACAACTACTAAGCCGAAACTTACAG 3009
647 LeSerAlaAsnValAspMetPheProLysValLeuGlnAspAlaMetLsn 663
3010 AGCATTTGCTACTTGTGATGAAATGATGATGCTCTGGAAGAA.....ACG 3053
664 LysArgAspGlnGlnLysArgLeuGlnLysAsnPheAlaAsnGlyMetLysG 680
3054 GAGCTCTCTATTAAGATCGCTGCGGAAAACTTGAAA.....ATGCATCA 3097
680 uMetLeuAlaProPheGlu.....HisLeuAlaArgLysAsn 693
3098 AACTACTTGAATTTAATTAAGAAAGCATTAAGATCGCTGAGCTGACA 3147
693 LysValCysProCysCysGlnArgAlaPheThrProAspGlnGlnLysP 709
3148 GCTGCTGTATTTGCTTCTGTAAGAAATTT.....GAAAAATGAATCG 3188
710 PheValLysLys.....GlnArgMetGlnAsnSerSerThrAlaGlnArg 724
3189 TTTAAAGTGAATTTGTACAGAACTAAACAAACAAACAGACGCAAAATTT 3238
724 gSerLysAlaLeuAlaMetGlnSerSerAsnAlaGlnAlaLeuPheGln 741
3239 TCAAAAACCTTTGAAAGACACCTGTCGCAAAACGAAAG.....GAAAT 3279
741 LLeuAspLysLeuArgThrLysThrAspAlaValLysLysLeu..... 755
3280 ATCTTCATAGTTTAAAGCTACTGAGAAACATATATATACACTAAATCT 3329
756 ValGlnGlnThrIleProLeuAlaGlnLysAsnLeuAsnGlnHisLeuAl 772
3330 ATAAATGAAAGATGATTAATTCACAGAAAGCTTGAAAAAGCTAAGCA 3379
772 AspGlnSerGlnLysAlaGlnAlaPheAspAspLeuGlnLysValLeu 789
3380 GGAACCAAAACCTTCCAAATCAAACTAGATGAATTTGGAAGTAGATTCGA 3429
789 LAnHisValGlnMetAspArgAspAlaValGlnAlaLeuLeuGlnProThr 805
3430 CAAATTTGAAGATGAGAAAGCTGAGTCTGAAATTCGTCCTTTG 3479
806 AspThrIleAspArgHisValHisGlnIleGlnGlnLeuValLysGlnVal 822
3480 .....ATTGAGAAAGTTTACTTACTTAAGAAAGCAACT 3511
822 GlnLysAspLeuGlnLysAlaLeuAspSerSerGlyArgGlyValLysSer 839
3512 TAAGGATCTTGAA.....AACAGTT 3531
839 euGlnGlnIleGlnLeuGlnLeuAsnPheLeuGlnArgThrArgAspThr 855
3532 CTAACACGATATCTGAGAGCTATGATTTATATACCAAGTGAAGATGCT 3581
856 LeuIleValGlnValAspAspLeuArgAspGlnHisArgMetLeuAsnG 872
3582 ...ATTCAAACTGTTGATGAATTAAGGACCAACAAAGAAAGATGAAGA 3628
872 uAspMet.....SerSerAlaGlnValArgThr 882

```

```

3629 TTTCTTAGCTGAGTGAAGAAACCATTTACAGATTTGCAATGAA..... 3674
882 LAsnAlaArgGlnGlnLysValLysAlaSerSerLysLeuGlnArgPhe 898
3675 .....AAAGATGAAAGGATGAGAGAAATTAAGATGATTAAGCTTG 3716
899 GlnLysSerGlnGlnGlnLeuValLeuLeuAlaGlnGlnLysGlnGln 915
3717 ATTAAGAAAGAAAGAA.....TTAACAGTTTCTGAAATTAATCAATC 3757
915 uIleValGlnLysLysLeuLeuGlnGlnLys..... 925
3758 TTTACACAGAAACAAATATGATGATTTCTTAAGCTCAAAAGCAAA 3807
926 .....LeuAspProLeuSerLysGlnLysGlnSerLeuGln 938
3808 ACATACAGATATCGATCTGAGTAAGAGATTAAGAGACAGCTATTT... 3854
939 GlnLysAsnAlaLeuLysGlnLysLeuAspGlnGlnLysGlnLeuAl 955
3855 .....ATTCTATTAACAAACAAAGATGAAAGCGCAAGTGTCTAGA 3898
955 ArgLysArgLys.....ArgGlnPheGlnGlnGlnL 965
3899 CAAACTAAATAATGACGATGATTCAGTACGCAACAAACAAACAGC 3948
965 euAspAlaLeuGlnLysArgLeu..... 971
3949 TTGCAATATTAATCGTTAATAGATTAATTCAGACCATTTATACGAA 3998
972 .....AsnMetLysLysLysGlnLysLysSerLysAsnGlnLys 986
3999 GTGCTGATTTTGAAGCTAAGGCTTC.....GATGATTTGCAAAACAA 4042
986 sLeuLysGlnLeuGlnGlnLysArgHisValLeuCysHisSerGlnLeuGln 1003
4043 AATTAATAAAGCTAGAA.....TTGACAAACCTCAAAAGCTGG 4080
1003 eCysMetAlaLysGlnGlnArgLysSerAlaGlnLeuAsnLysSerLys 1019
4081 AACTTAAGACACAGCTGATCTCAAGTCAAAATGAAGTAAACGAAAGAA 4130
1020 GlnLeuLeuGlnGlnGlnGlnLysLysArgAsnIleAspAspAsnLe 1036
4131 AGAAAACTGCTGACTCAATATGAGAGAAACCTGGAAGCAAACT 4180
1036 uLysLysArgLysThrLysAlaAspValGlnIleuThrArgAspIleG 1053
4181 AGAAGTATTTGAATTTGAATTCACAACTTCACACACTTAATCTGAGATTA 4230
1053 LysLeuGlnGlnLysArgLeuLeuSerLysLeuSerLysAlaIleGln 1069
4231 GTAGATTAGAC.....GTCAAAATGACAGAA 4256
1070 AlaAspLeuLysArgHisSerGlnGlnLysGlnLysLeuAsnSerGlnP 1086
4257 GCTAAGAGGACCAATATCAAGAGAAATCACTAAGCAATCAAAAGAT 4306
1086 eAsnArgTrpGln.....GlyThrLeuSerValT 1096
4307 TGAAGAGCTGACCTGCAAAACGCAAGTAAATTAAGTCAAAATGAACAT 4356
1096 LysGlnSerAsnIleSerLysHisLysGlnGlnLysLysLeuSerGlnLys 1112
4357 TACAGAAATCAATAGATTCATGACTCATCACTGCGG...ACGATATAC 4403
1113 LysAspIleGlnLysArgGlyThrAsnGlnPheLeuGlnLysLysThr 1129
4404 AAGATATCGAGAAATATTAACATGAAGATGGGTGATTAACAGACAAAG 4453
1129 rGlnMetAlaAsnLysAspLeuAspArgGlyThrThrAlaLeuAspLys 1146
4454 ATCTTTGTTACTAGCAGCATAGATGCTATTCAAAGCATTAAGTAGTG 4503

```

```

1146 146mefarphhisermetymetgucguil14easny11elle 1162
1147 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4504 CTAATATGAAATATCATGATGATTGAAAGCAAGATATTAACAGGATANTT 4553
1163 lvsGluLeuTrpGlnInThrTyrArgLyGlnaspIleasPyrIleSe 1179
1164 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4554 GATGATATATGGAACGATACGATATGCTACTGATATATGACACACACAA 4603
1179 r11easnsrAsp.....SergLUGLYALGLYThrArgSerTyrSert 1194
1180 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4604 AATTGATCATCGATGAGTATGATAGTACGCTTAAGGAGAAATCAACACACT 4653
1194 yvArzVAlVAlmetGlnThrGlyAspAlaGluLeuGlnMetArgGlyArg 1210
1195 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4654 ATCGGTTGTATGATGACAAAGCAGATGTGTAATGATGATGAGAGACGA 4703
1211 cySseRAlaGlyGlnLySValLeuAlaSerLeuIleIleArgLeuAlaLe 1227
1212 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4704 TCTTCGGGGGCGTCAAAAAGTTCTTGCATCGATTTATTATCAGCGTACGCT 4753
1227 uAlaGluThrPheCysLeuAsnCysGlyTleLeuAlaLeuAspGluPOT 1244
1228 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4754 ATCTGAACCATTTGGCGCAATTTGGCGGCTAATTCACATGACACACCTTA 4803
1244 hTThrAspLeuAspGlyProAsnAlaGluSerLeuAlaAlaLeuLeu 1260
1245 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4804 CCACCAATTTAGATGACAGAAACATCGAAGAGTTTATGCAAAATCTTTAC 4853
1261 ArgIleMetGluAlaArgLySGLyGlnGluAsnPhgGlnLeuIleVal11 1277
1262 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4854 AATATATATTAATATGAGAAAGCATCAACATAAAGATTTCCAGTAACGTCAT 4903
1277 eThrHisAspGluArgPheAlaHisLeuIleGlyGlnArgGlnLeuAlaG 1294
1278 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4904 CACTCATGTGTGAGAAATTTTGGCTCATATGAATGCACGGCATTTACTCG 4953
1294 lUlyTYTTrArGvAlSerLyAspGluAsnGlnHisSerTleIleGlu 1310
1295 ::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4954 ATCACTTCTTCAAAATCAAGAGAGATGATAGCAAAAATCCAAATTGAG 5003
1311 SerGlnGluIle 1314
1312 :|||||
5004 TGGGTGCGATAT 5015

seq_name: qb_D13:SCCHXIVL

seq_documentation_block:
LOCUS SCCHXIVL 33016 bp DNA PLN 30-JUL-1997
DEFINITION S.cerevisiae DNA region from chromosome XIV, left arm.
ACCESSION X66722
VERSION X66722.1 GI:1255956
KEYWORDS ATX1 gene; MP43 gene; NRD1 gene; ORC5 gene; RAD50 gene; RPA49
gene; SIR3 gene; su11 gene.
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 33016)
AUTHORS Sen-Gupta,M., Guldeher,U., Beinhauer,J., Fiedler,T. and
Hegemann,J.H.
JOURNAL Direct Submission
MEDLINE Hegemann,J.H.
REFERENCE Submitted (21-MAR-1996) J.H. Hegemann, Institute of Micro &
MOLECULAR BIOLOGY, Justus-Liebig-Univ. Giessen, Frankfurt Str.
107, D- 35392 Giessen, FRG
FEATURES
source 1..33016
Location/Qualifiers

```

gene  
/organism="Saccharomyces cerevisiae"  
/strain="S288C"  
/sub\_strain="F11679"  
/db\_xref="taxon:4932"  
/chromosome="XIV"  
/clone="cosmid 14-4c pou"  
/map="132kb, left arm"  
223..1662  
/gene="ORC5"  
223..1662  
/gene="ORC5"  
/note="ORF N0834"  
/codon\_start=1  
/protein\_id="CA65483.1"  
/db\_xref="GI:1253937"  
/translation="MNVTPEVAFREYOTNCIASYISADPIDTPSNTILQYSGTGKTN  
YTLKFKFNPNPMLHAVMLEPVELVSMKRLQAIATVQYKLTLYPNIPPTDYDPLQV  
BEPLLITKLPHNIHFOYSEIOERKCLPLILDGPSLODLAAFNKTKIKNELPKQSS  
KINIKFIYMETSTSLQKSTHCIFTPVAPRYNDVSTIIVMSCEIGMDSGLRRKR  
IIIEQDTLDQDFQNVAAFNTHLVQAFHSTGNDIPALNDLIDFKMPKVSITKE  
NFEEDPATKSAIKFLSTDDMLSENGGESAITTDRLDLENSQYDLSITSKLLIA  
SYCSILERYDASIFSKRTYICQRAAYGRKKKEVPRYIQSLPAIEBLAIFQAPAL  
IPIPGKAGSGLSALREESLKANIIEVQNLSELHTLKLIATNMKNIDYLSPKVRM  
KNVPMETIKETSESHVFNISDYFSDIHE"  
complement(1982..2578)  
/note="ORF N0838"  
/codon\_start=1  
/protein\_id="CA65484.1"  
/db\_xref="GI:1253958"  
/db\_xref="SWISS-PROT:P53846"  
/translation="WNRNFIKKKKNFPSKNSHLSYLSKKKVKITGOIKKDPNLLN  
LEEQYQSEFLLEGONENIKOSFLEGRQGLDVGQRFLLQMGGLDVESYGLHSP  
TEKNIHTRITLKGILKNMNDSEYMEERVYLIRKKNFRTILTHLVKDKRTPPV  
TEVEEDVSRAIAGEIRGVEVEDIAKKKTKONQASW"  
complement(2767..2988)  
/gene="Arx1"  
complement(2767..2988)  
/gene="Arx1"  
complement(2767..2988)  
/note="ORF N0840"  
/codon\_start=1  
/protein\_id="CA65485.1"  
/db\_xref="GI:1253959"  
/db\_xref="SWISS-PROT:P38636"  
/translation="MAEIKHYQFNVMVSCGSGAVNKKVILKLEPDVSKIDISLEKQD  
VVVYTLIPDYFLIEIKIKTKGKEVRSKQI"  
complement(3232..5596)  
/note="ORF N0842"  
/codon\_start=1  
/protein\_id="CA65486.1"  
/db\_xref="GI:1253960"  
/db\_xref="SWISS-PROT:P53847"  
/translation="MESLEPNKGEIIRFELNDPLILKDKSRKNSGELELDSLOLQ  
EALIANELINDLTKLPTNLIREKVTNINILEENCYSLQSIKKKMNNAAYIKOSF  
NVAQISITYVDLHLELVSTLKILITNPFKTTENSIOPTPYVNGKDVHIEIDTFM  
DTPVQOYFPRGSLNDQAMPILDMTSDAQEOVRAKLIMKEYNMLSTVSMINSFV  
IGSEKTEFENKNIITVFSKSSSHGQCVSTVLTSEAVCDMLDGLAFRDKTILSYEL  
GPLENTEFPRVKNNAISILLESIDPLKNIIVSVINNKLRPLVAKSEVYNMTHSGKEIO  
DILMNKOYLVNLLDKVLESHISSEIRSFEDPKRSQMLLEYELTSTNTMSKIKKE  
NDSVDONKEKLNNAYSKDDMWREVEDDDADAMQEDIVNIDDEBKTNQKESPEER  
EEBAMDEAALIDENIDDAIDSENGEKHLAHADGSLDKOHIEVQLPKPLAISQNFSS  
SPADSHVQKMYRIYQLQITSYAMACFANSHNNQCYVMRYLIEDEKLYIKEL  
TNNLTLETKMAYRIYQILRIHQLTPEFENENRPNMDATIEKLPYLITKEIVPRLQII  
RGEGRGKRYFLSENLNDCVTKREIKLMOYISEVSENMIGELVSLVNMVQIQLAAKE  
PSYKKREKFAIMGKFLPLHLKEIMENYNGDFYLAFTDELIONTELLFADTPLRRNA  
IDDIYIIRGTALDD"  
complement(5754..9443)  
/gene="SIP3"  
complement(5754..9443)  
/gene="SIP3"  
/note="ORF N0844"  
/codon\_start=1  
/protein\_id="CA65487.1"



```
20927 CTCGTGATTTACGCACAGACAGTCCCTGATCTGGAGTTCCTCAAGG 20976
      ::::::::::: ::::::::::: :::::::::::
151 VALLEUAGUASNAVALIephEVALHISGLNASPGIUSERSATRPProle 167
      ::::::::::: ::::::::::: :::::::::::
20977 ATCTTGAAATACGTGATATCTGTGCACCAAGAGACAGCTGTGCACCT 21026
      ::::::::::: ::::::::::: :::::::::::
167 UGLNASPROSERThrLeuLYSLySLySPheASpRIepheserAlat 184
      ::::::::::: ::::::::::: :::::::::::
21027 GAGTGAACCATCCAACTTTAAAAAGAGTTGATGATAAATTTTCCAGCCA 21076
      ::::::::::: ::::::::::: :::::::::::
184 hrArgTYThrLYSAlaleuGLUVALIleLYSLySerHISLYSASpGIN 200
      ::::::::::: ::::::::::: :::::::::::
21077 TGAATTTACAAAGGCTCTAGCAATTTAAAAATCCATAAGAAAGCATG 21126
      ::::::::::: ::::::::::: :::::::::::
201 MetGLINuILELYSThrPheArgLEUySLyGLUASnLEUGlnThrVA 217
      ::::::::::: ::::::::::: :::::::::::
21127 TCCCGATATCAAGTATTATAAGCAATCAGTAGAAGCACTTAAGTGG.. 21174
      ::::::::::: ::::::::::: :::::::::::
217 LLYSASpGINAlaHISLYSLeuArgGLUASnIleAGINASpGINLU 234
      ::::::::::: ::::::::::: :::::::::::
21175 ..:::..GATMAAGACA 21184
      ::::::::::: ::::::::::: :::::::::::
224 yASerASpAlaserLYSserGINMetGLUlnLEuLYSGLUySLIEcyS 250
      ::::::::::: ::::::::::: :::::::::::
21185 GATCGAAAGCAATGAGCTCAATATCCACCAATTCGAAACTAAGATAGAT 21234
      ::::::::::: ::::::::::: :::::::::::
251 GLYThrGLUArgLUileuGLINMetGLUThrSerLEUASpGIUleUAR 267
      ::::::::::: ::::::::::: :::::::::::
21235 CAGTATATGAAAGAGTGTCTGAATTCATCTCAATGATGAATGAATC.. 21282
      ::::::::::: ::::::::::: :::::::::::
267 gArgLEUGlnGLYlnIleASpRIeLYSAlaThrGLUArgSerThrLeuL 284
      ::::::::::: ::::::::::: :::::::::::
21283 ..:::..ACTGGAATTCGGATMAAC 21301
      ::::::::::: ::::::::::: :::::::::::
284 eu.....ThrGLINHISGLUySLyLeuAlaIleuSerGINLU 297
      ::::::::::: ::::::::::: :::::::::::
21302 TGTTTAAATCTNATCAAGATTTCCAAAAATACTATCCAAAGTAGAGAAC 21351
      ::::::::::: ::::::::::: :::::::::::
298 ASnGLUASpThrASpGIUleuLEUmetGLUThrPGLINThrLYSPheGLU 314
      ::::::::::: ::::::::::: :::::::::::
21352 TTGAAATAATCAAAACTCTCATCTAGTGTATCAAGTGAAGATGTGCAAA 21401
      ::::::::::: ::::::::::: :::::::::::
314 uArgIleAlaLeuLeuGLU.....T 321
      ::::::::::: ::::::::::: :::::::::::
21402 CTCATTTGATATTCGATTTATCGAAGCCAGATTTCAAAACTATTTG 21451
      ::::::::::: ::::::::::: :::::::::::
321 hrLYSIIeserLYSLeuVALArgASP.....MetASpASP 332
      ::::::::::: ::::::::::: :::::::::::
21452 CAAACTTCTCTAAAGCTTGATGATMAAGATATCAGTTGAGAGATTGG 21501
      ::::::::::: ::::::::::: :::::::::::
333 GIUAlaserTYrSerSerValLeuSerLYSGLINASer.....GLUle 347
      ::::::::::: ::::::::::: :::::::::::
21502 GAACCGATATTTCAGATTGGAAGCCGTGCTCATCTCTGCAGAGCCT 21551
      ::::::::::: ::::::::::: :::::::::::
347 uThrHISGLUileGLYLYSLyLeuGLINAlaGLUAlaASpAlaHISLeuThM 364
      ::::::::::: ::::::::::: :::::::::::
21552 TTCTAATTCCTCATCAGCGCGGAGGTGAGTTAGAACT..... 21591
      ::::::::::: ::::::::::: :::::::::::
364 eLYSHISGLUArgASpSerASpRIeLYSASnIleCYSThrLYSHISASn 380
      ::::::::::: ::::::::::: :::::::::::
21592 ..:::..GCTAAGAGACATCGAAAGAACAGAAATCACCCTTTCATCG 21633
      ::::::::::: ::::::::::: :::::::::::
381 LeuGLYProVALProGLUHISProPheThrASpVALAlaMetASnle 397
      ::::::::::: ::::::::::: :::::::::::
21634 CTAAAGAAAGCCTTCCACATAGTTC.....CAAGGCT 21668
      ::::::::::: ::::::::::: :::::::::::
397 uThrASnArgIleLYSAlaArgLEUserSerLEUGINASpRIeLeuA 414
      ::::::::::: ::::::::::: :::::::::::
21669 TTCAATATATGAAACAGATGATGCTCAATAATTCAGAAATGTCAAC 21718
      ::::::::::: ::::::::::: :::::::::::
414 sp...LYSLyLYSserASnGLUASpGINLeuASpAlaLeuTRpLYSHIS 429
      ::::::::::: ::::::::::: :::::::::::
21719 AGTCAAGCCTTTCATCTCTCAGATTTAAGTACATAAGTCAATTT 21768
      ::::::::::: ::::::::::: :::::::::::
430 TYrLEUySLIleASnAlaArgTYrSerGLUVALASpGIYlnIleGLINse 446
      ::::::::::: ::::::::::: :::::::::::
21769 GCAAAGGATATTCACCTAAAGAAACAAATTTATCCGACTGATTAATTC 21818
      ::::::::::: ::::::::::: :::::::::::
446 rLYSIIleGLUserMetSerGLYIleuArgArgArgLYSASpLYSGLU 463
      ::::::::::: ::::::::::: :::::::::::
21819 AATTAACCGTCGATTCACAGATTTTGGAAATTAATAAAGGACATCTCA 21868
      ::::::::::: ::::::::::: :::::::::::
463 yS.....GIUArgASpAlaAlaGLUVALGIUleuSerLYSPhe 475
      ::::::::::: ::::::::::: :::::::::::
21869 AATTAATCATGACTCTGAAGAGCTTGCAGAAATTTGAATCAATTTAA 21918
      ::::::::::: ::::::::::: :::::::::::
476 ASnLeuSerArgIleASpGIUArgGLUArgHISmetGLIN..IleGLUVA 491
      ::::::::::: ::::::::::: :::::::::::
21919 AGTTTCTACACCAAGATAGCCTCAATCATGTGAACGTGAAGAACTTAAAC 21968
      ::::::::::: ::::::::::: :::::::::::
491 lGIUArgLYSThrLeuAlaLeuGLYGLUArgASP.....T 503
      ::::::::::: ::::::::::: :::::::::::
21969 ATRCAAAAGAAAGCTCAATCTTGGAAAGTGAAGATTTATTAACCAAA 22018
      ::::::::::: ::::::::::: :::::::::::
503 yRASpSerIleIleSerGINLYSArgThrGLUVALySerLEUGln 519
      ::::::::::: ::::::::::: :::::::::::
22019 TGAAACGAAATTAAGAGAGAAATAATAGATGATTTCTTAGAGAT 22068
      ::::::::::: ::::::::::: :::::::::::
520 LYSIIleLYSVALleuLeuArgGLUySLASpRIleASnArgASnAlas 536
      ::::::::::: ::::::::::: :::::::::::
22069 CAATATGAGAAATTTCCAGATCGCATTAAGAAACCAACCAAGCGGA 22118
      ::::::::::: ::::::::::: :::::::::::
536 pGIUArgVALySLyLeuGLYleuLYSASpAlaLeuGLUserLYSVA 553
      ::::::::::: ::::::::::: :::::::::::
22119 CTTATACGCCCAACTAGAGCTATTAAGAGATTAATTAATTAATG 22168
      ::::::::::: ::::::::::: :::::::::::
553 sPLySLyASnGLUileVALASn.....GLUHISLYSASpLYSILEyS 567
      ::::::::::: ::::::::::: :::::::::::
22169 ACGAGTTACAAAGATTTACGGAATAATTCAGAAATGATCTCGAATTAAG 22218
      ::::::::::: ::::::::::: :::::::::::
568 LYSVALleuArgGLYArgASnProPheGLUyS...ASpMetLYSGL 583
      ::::::::::: ::::::::::: :::::::::::
22219 CAAGTATTTCCCTCACCCAGGAATTTCAAGAGCGATTTAGAAATGGA 22268
      ::::::::::: ::::::::::: :::::::::::
583 uIleASnGLINAlaPheTrpProVALASpLYSGLUyThrASnGLUleuAys 600
      ::::::::::: ::::::::::: :::::::::::
22269 TTTTCAGAAATTAATTCATCAATATGCAAAAGATATAGCGATCAACAA 22318
      ::::::::::: ::::::::::: :::::::::::
600 eLYSserGINLUAlaGLUlnGLUleuLYSPheThrGLINserLYSVAL 616
      ::::::::::: ::::::::::: :::::::::::
22319 AANAAGTCACGAAATTTGATAG.....AGATAT 22347
      ::::::::::: ::::::::::: :::::::::::
617 ThrASpAlaArgGLUlnLeuThrLYSLeuArgArgASpMetASpAlaLY 633
      ::::::::::: ::::::::::: :::::::::::
22348 ACAAGCGCTTTGTACAAATTTGAACACCATTTGAAGAGATTTACAGATTA 22397
      ::::::::::: ::::::::::: :::::::::::
633 SARArgPheLeuASpSerLYSLeuGLINserIle.....LeuGLINI 647
      ::::::::::: ::::::::::: :::::::::::
22398 TCAGAGTGTGAAGAAAGAAAGTATACACTACTAAGCGAAACTTAACAG 22447
      ::::::::::: ::::::::::: :::::::::::
647 IeserAlaASnVALASpMetPheProLYSVALleuGLINASpAlaMetASn 663
      ::::::::::: ::::::::::: :::::::::::
22448 AGGATGTGCTATTTGATGATTAATGATGCTCTGGAAGAA.....ACG 22491
      ::::::::::: ::::::::::: :::::::::::
664 LYSArgASpGLUlnLYSArgLEUGINASnPheAlaASnGLYMetArgGL 680
      ::::::::::: ::::::::::: :::::::::::
22492 GAGCTCTCTAATAAGACTCGCTGGAAGAACTTGAAA.....ATGCATCA 22535
      ::::::::::: ::::::::::: :::::::::::
680 uMetLeuAlaProPheGLU.....HISLeuAlaArgLYSASnH 693
      ::::::::::: ::::::::::: :::::::::::
22536 AACTACTTTAGAAATTAATTAAGAAACATTAAGAGATCGCTGAACGTGADA 22585
      ::::::::::: ::::::::::: :::::::::::
693 IseVALySProCYScysGLUArgAlaPheThrProASpGINLUASpGIN 709
      ::::::::::: ::::::::::: :::::::::::
22586 GCTGCTGTTATTTGCTTCTAGAAAGTT.....GAAATGAATGCG 22626
      ::::::::::: ::::::::::: :::::::::::
```





```
24292 ATATATATATATATGAGAGACATCAAAAGATTTCAGCTAATCTGAT 24341
1277 eThRHisAspGluArgPheAlaHisLeuIleGlyGlnArgGlnLeuAlaG 1294
24342 CACTCATGATGAGAAATTTTGGTCATATGAATGCAGCGCATTTACTG 24391
1294 LuLysTyrTyrArgValSerLysAspGluAsnGlnHisSerIleIleGlu 1310
24392 ATCACTTCTTCAAGTCAAGAGAGATGATAGGCCAAATAACACAAATTGAG 24441
1311 SerGlnGluIle 1314
24442 TGGGTGATATT 24453
```



adapter of sequence [5'  
GAGGAGAGATTCGAGATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved  
with BamHI and XhoI. Vector: a modified pBluescript KS(+) after  
bulk excision from Lambda Flc I. Cloning sites, 5' end: SalI; 3'  
end: BamHI. Host: DH10B.

# FEATURES

source Location/Qualifiers

1..907  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGI:109292"  
/db\_xref="MGI:1909537"  
/clone="5830450E06"  
/sex="male"  
/tissue\_type="thymus"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
265..>907  
/note="putative"  
/codon\_start=1  
/protein\_id="BAB31030.1"  
/db\_xref="GI:12857523"  
/translation="MSRIKMSIILGVRSPGIEDKDKQIISFSPITLIVGPNAGKTT  
ITECKYICGDPGPPGTGNTFVHDPRVAQEDVRAQRIQEDVNGEMVAQVRSMLC  
SKRNKTEKRTIEGVTTRMKHGEKYSLSKCAEIDREMTSCIGSKSVINNYIFHOE  
DSNWPUSSEKALKQKDELFSKRTIKALDTLRQVQKQKVECTELATLK"

BASE COUNT 247 a 199 c 242 g 219 t

ORIGIN

alignment\_scores:  
Quality: 580.50 Length: 217  
Ratio: 3.243 Gaps: 3  
Percent Similarity: 82.488 Percent Identity: 52.074

alignment\_block:  
US-09-538-396-2 x AK018001 ..

Align seg 1/1 to: AK018001 from: 1 to: 907

1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPhe.... 15  
|||||.....  
265 ATGTCCCGGATCGAAGATGAGCATTTGGCGCGCAAAATTTGGAT 314  
16 AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeu 32  
:.....  
315 AGAGATTAAGATTAAGCATTAATCTCTTTCAGCCCTCAACAATTT 364  
32 LevalGlyProAsnGlyAlaGlyLysThrIleIleGlyCysLeuLys 48  
:.....  
365 TGGTTGGACCAATGGGCGGGAAGACGACCATTAATGCTTAAG 414  
49 LeuSerCysThrGlyLysLeuProProAsnSerArgSerGlyHisThrPh 65  
|||||.....  
415 TATATTATTCTGGGATTTCCCTCTCGAAACCAA...GAAATACATT 461  
65 eValHisAspProLysValAlaGlyLysThrGlyLysGlyLeuIleL 82  
|||||.....  
462 TGTTCATGATCCCAAGTTGCTCAAGAAACAGACGCGGCGCCAGATTTC 511  
82 yLeuLeuArgPheLysThrAlaAlaGlyLysAspValAlaCysIleArgSer 98  
:.....  
512 GCCTCCAGATTCGAGATGTGAATGAGAGATGCTAGCTGTCCAAGGCTCT 561  
99 PheGlnLeuThrGlnLysAlaSerLysMetGlnPheLysAlaIleGluSe 115  
:.....  
562 ATGCTTTGCACTCAGAAATAATTAATAACAGATTTAAACCTTGGAAG 611  
115 tValLeuGlnThrIleAsnProHisThrGlyLysValCysLeuSerT 132  
:.....  
612 AOTCATTTACTAGATGAAG.....CATGGTGAATAAAGTCAAGTTCACCT 655  
132 yTArgCysAlaAspMetAspArgGluIleProAlaLeuMetGlyValSer 148

seq\_name: gb\_est45:AM275863

seq\_documentation\_block:

LOCUS AM275863 596 bp mRNA EST 03-JAN-2000  
DEFINITION xq40a12.x1 NCI-CGAP\_Lu28 Homo sapiens cDNA clone IMAGE:2753086 3'  
similar to TR:092878 Q92878 RAD50. ; mRNA sequence.

ACCESSION AM275863  
VERSION AM275863.1 GI:6662893

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 596)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: [cgapsb@emall.nih.gov](mailto:cgapsb@emall.nih.gov)

COMMENT Tissue Procurement: Chris Moskalko, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life  
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.lnl.gov/bdrrp/image/image.html](http://www.bio.lnl.gov/bdrrp/image/image.html)

Seq primer: -40UP from Gibco  
High quality sequence stop: 403.

Location/Qualifiers

1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="2753086"  
/clone\_lib="NCI-CGAP\_Lu28"  
/tissue\_type="two pooled squamous cell carcinomas"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: PCMV-SPORT6; Site: 1; Salt:  
Site: 2; NOTI: Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies."

# FEATURES

source Location/Qualifiers

1..154  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="2753086"  
/clone\_lib="NCI-CGAP\_Lu28"  
/tissue\_type="two pooled squamous cell carcinomas"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: PCMV-SPORT6; Site: 1; Salt:  
Site: 2; NOTI: Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies."

BASE COUNT 154 a 134 c 108 g 200 t

ORIGIN

alignment\_scores:  
Quality: 570.50 Length: 175  
Ratio: 3.803 Gaps: 1  
Percent Similarity: 85.714 Percent Identity: 64.571

Alignment block:  
US-09-538-396-2 x AM275863/rev

Align seg 1/1 to reverse of: AM275863 from: 1 to: 596

```

1140 TyrThrAlaLeuAspLysAlaLeuMetArgPheHisSerMetLysMetG1 1156
|||||
596 TATAAGCTTCGGACCAACATATGAAATTTTCCACTATGAAAATGGA 547
1156 uGluIleAsnLysIleIleLysGluLeuTrpGlnGlnThrTyArgLysG 1173
546 AGAATCATATAAATAATACGTGACCTGGCGAAGTACCTATCGGGAC 497
1173 LnaSpLleAspTyrIleSerIleAsnSerAspSerGlu..... 1185
496 AAGATATTGAATACATAGAAATACGGTCTGATGCGGATGAAAATGATCA 447
1186 ...GlyAlaGlyThrArgSerTyrSerTyrArgValValMetGlnThrG1 1201
446 GTTCTGTGATAAAGCGGCAATATACACCGACGTCGATCGTCAAGG 397
1201 YAspAlaGluLeuGluMetArgLysArgCysSerAlaGlyGlnLysValL 1218
396 AGACACACGCTGGATATGCGAGACGATGCGAGTCTGACAAAGGTAT 347
1218 euAlaSerLeuIleIleArgLeuAlaLeuAlaGluThrPheCysLeuAsn 1234
346 TAGCTCTCATCATCATCGCTGCGCCGCTGCGAAGCGTTCGCTCAAC 297
1235 CysGlyIleLeuAlaLeuAspGluProThrThrAsnLeuAspGlyProAs 1251
296 TGTGCAATCATTTGCTGTGATGAGCCACACAAATCTTGACCGAGAAA 247
1251 nAlaGluSerLeuAlaAlaLeuLeuArgIleMetGluAlaArgLysG 1268
246 CATGCAATCTCTTGACATGCTCGTGTGAGATATATAAAGTCCGTCAC 197
1268 LylGlnGluAsnPheGlnLeuIleValIleThrHisAspGluArgPheAla 1284
196 AGCAGCGTAACCTCCAGCTTCGTGTAATCACAATGATGAAATTTTGG 147
1285 HisIleuIleGlyLysArgLysLeuAlaGluLysTyrTyrArgValSerLys 1301
146 GAGCTTTTGAAGACGTTCTGAAATATGAGGAGAAATCTTACAGATTA 97
1301 sAspGluAsnGlnHisSerIleIle 1309
96 GAACATCGATCATCGTCTCAGAGATT 72

```

seq\_name: gb\_gss23:AZ216324

seq\_documentation\_block:

LOCUS AZ216324 597 bp DNA GSS 09-JUN-2000  
DEFINITION Sheared DNA-8085.TF Sheared DNA Trypanosoma brucei genomic clone  
ACCESSION AZ216324  
VERSION AZ216324.1 GI:8434124  
KEYWORDS GSS.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei.  
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE  
AUTHORS 1 (bases 1 to 597)  
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Sub, E., Malek, J., Fujii, C.,  
Gerrard, C., Leech, V., de Jong, P., Willu, E., Melville, S., Donelson, J.,  
Fraser, C. and Adams, M.

TITLE Determination of clone end sequences from Trypanosoma brucei GUTat  
10.1 sheared DNA library  
JOURNAL Unpublished (1999)  
COMMENT Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA

# FEATURES

## source

Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
DNA library constructed at TIGR. Clones will be available for  
distribution through Research Genetics, Alabama, USA. Sheared DNA  
end sequences search page: <http://www.tigr.org/tldb/mdb/tldb/>.  
Seq primer: M13-Forward  
Class: shotgun.

## Location/Qualifiers

1..597  
/organism="Trypanosoma brucei"  
/strain="TREU927/4 GUTat 10.1"  
/db\_xref="taxon:5691"  
/clone="Sheared DNA-8085"  
/note="Vector: pUC18; Site 1: SmaI; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically  
sheared to give a tight size distribution (approx 2 kb).  
The v + l method used for the library construction is  
described in detail in Smith, H.O. and Venter, J.C.  
(Making small insert libraries for whole genome shotgun  
sequencing projects. In Genome Sequencing: A Practical  
Approach, eds. M. Vaudin and B. Barrel, Oxford University  
Press, 1999)."

BASE COUNT 160 a 134 c 156 g 147 t  
ORIGIN

## alignment\_scores:

Quality: 561.00 Length: 194  
Ratio: 3.380 Gaps: 2  
Percent Similarity: 85.567 Percent Identity: 54.639

alignment\_block:  
US-09-538-396-2 x AZ216324

Align seg 1/1 to: AZ216324 from: 1 to: 597

```

1112 TYRYSASPILLEGILULYSARGTYTRHANSGLNPHLEUGLNULEULYS 1128
|||||
1 TCAAAAGATATGAAAAGCGCTACCGACGACCTTCCTGAAGTCCAGAC 50
1128 rThrgLumetaLAsnLysAspLeuAspArgTyrThrAlaLeuAspL 1145
|||||
51 TACTGAATTTGCGGTTTCGATGTGAAAAATATATATCGACGCGTGAAA 100
1145 yAlaLeuMetArgPheHisSerMetLysMetGluGluIleAsnLysIle 1161
|||||
101 AGCAGCTGACAGACTACACCAAGAGAAATGCGGATCAATCAATCATC 150
1162 IleLysGluLeuTrpGlnGlnThrTyrArgLysGlnAspTyrIle 1178
|||||
151 TTGGCTGACCTTTGGCGACACACCTTACAGGAGAGCATATGACACAT 200
1178 eSerIleAsnSerAspSerGluGlyAlaGlyThr.....ArgSerT 1192
|||||
201 TGAATCTCGCTCCGAGATGATGTCACTCACTACAGCAGAGCGAGTT 250
1192 ySerTyrArgValValMetGlnThrGlyAspAlaGluLeuGluMetArg 1208
|||||
251 ACAGCTATCGTGTGATGAAAGGGGGAACGTAAATGGCATGCGG 300
1209 GlyArgCysSerAlaGlyGlnLysValLeuAlaSerLeuIleArgPhe 1225
|||||
301 GGCGCTGACAGTCGGGCTCAAAAGTACTTGCCTCTCTCATCGCACT 350
1225 uAlaLeuAlaGluThrPheCysLeuAsnCysGlyIleLeuAlaLeuAspG 1242
|||||
351 TCGCTAAGAGAACCTTTGTGTGACTGCGGAATCTTGATGATGATG 400

```

1242 luprothrtthraenleuaspjlyproasna1aglusertleuata1a1a 1258  
 401 AGCGGACAGAGAACTTGAGAGATAGACGGGCTTCTTGGGAGTCG 450  
 1259 leuenuargilemetglua1aarglysglynglunspheglnleu1 1275  
 451 CTACGATGCTTATTGACACTGACGAGTGAACATTTTCAGCTTAT 500  
 1275 eallleththraenleuaspjlyproasna1aglusertleuata1a1a 1292  
 501 TGTATTACTACAGATAGACACTTGTCCGCCGCTAGAGGCCAGCAC 550  
 1292 eu1ag1ulys1tyr1arg1valser1lyasp 1302  
 551 TG...GACTTTTATTACATTCATAGAT 579  
 seq\_name: gb\_est41:AM029964

## seq\_documentation\_block:

LOCUS AM029964 561 bp mRNA EST 15-SEP-1999  
 DEFINITION EST073219 tomato callus, TAMU Lycopersicon esculentum cDNA clone  
 CLEC12L16, mRNA sequence.

ACCESSION AM029964 GI:5888720

## KEYWORDS

EST.

## SOURCE

tomato.

## ORGANISM

Lycopersicon esculentum

## REFERENCE

1 (bases 1 to 561)

## AUTHORS

Alcala, J., Vrebalov, J., White, R., Matero, A. L., Vision, T., Holt, I. E.,  
 Liang, F., Dpton, J., Craven, M. B., Bowman, C. L., Ahn, S., Konning,  
 C. M., Fraser, C. M., Martin, G. B., Tanksley, S. D., and Giovannoni, J.

## TITLE

Generation of ESTs from tomato callus tissue

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: David Frisch  
 Clemson University Genomics Institute  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 4366  
 Fax: 864 656 4293  
 Email: dfri@chelemson.edu

## FEATURES

5 prime sequence.

## source

location/Qualifiers

## 1..561

/organism="Lycopersicon esculentum"

## /cultivar="TA496"

/db\_xref="taxon:4081"

## /clone="CLEC12L16"

/clone\_1lb="tomato callus, TAMU"

## /tissue\_type="callus"

/dev\_stage="25-40 days old"

## /lab\_host="X11-Blue MRF"

/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; supplier: Giovannoni laboratory, CLEC - Cotyledons  
 of seedlings 7-10 days post-germination were excised, cut  
 at both ends and placed on MS medium with no selection.  
 Mixed callus was harvested at 25 and 40 days and included  
 undifferentiated masses. Tomato Callus Est Library"

## BASE COUNT

207 a 86 c 125 g 143 t

## ORIGIN

185

## alignment\_scores:

Quality: 551.00  
 Ratio: 3.380  
 Percent Similarity: 88.108  
 Percent Identity: 57.297

## alignment\_block:

US-09-538-396-2 x AM029964

Align seg 1/1 to: AM029964 from: 1 to: 561

956 GUARGLYSARGGLUPHEGNGINGLULEUASPALALEUARGLEUAS 972  
 3 GAATTAAGAAGAAATATATATATATATATATATATATATATATAT 52  
 972 nmetlysllelysglytyrleuaspserlyllysaanglyllyleuysg 989  
 53 CTCGAAATTAAGAATATATATATATATATATATATATATATATAT 102  
 989 lueugnglyarghlsvalleucyshtsserlgnleugnginsercysmet 1005  
 103 AACCTGAGGAGCAAGGAGGCTGTCAAGATCTCAACCTCAAGTTGGAG 152  
 1006 A1Aysgnglnarglylleser1ag1ulys1tyr1arg1valser1lyasp 1022  
 153 TCCAGAAAGATGCGATTTTGGCTGAAGTCAAGAAAGTAAAGCATTTAAT 202  
 1022 uginglynglyluleuysargasn1leaspaspasnleuys1tyr 1039  
 203 GGGGAACAGAGATAGTTGAGAGTAACTTGAAGATTAAGTAACTATC 252  
 1039 rgllysthrlysa1aaspvalgluglneuthrargaspillegluserleu 1055  
 253 GTAAATTAAGTGTGAAGTGTGAGTCACTCATGATGATTAATTAATCTG 302  
 1056 GLUGLARGLEUASERILEGLYSERLEUSER1ALLEG1UALAASPLE 1072  
 303 GAAGATTAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 352  
 1072 ulysarghls1sergnglyluleuysargasn1leaspasnleuys1tyr 1089  
 353 TAAAACTTCAACAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402  
 1089 rpglnglythrleuservaltytyglinsersn1leserlyshlysgln 1105  
 403 GTCATGCGACCTTATCTGTATCAAGCAATATCTCAAGCAAGAGT 452  
 1106 GLULEUYSLEUSERGNTYR1LYSASP1LEG1ULYSARGTYRTHRAENG 1122  
 453 GACCTTAAGCAGGACATATCAAGATATATGCAAGAGCTATTTTGATCA 502  
 1122 nphleuglneulys1thrglmetala1asnlysa1leuasp1argt 1139  
 503 GCTAATCCAGCTTAAGACTAGGAGATGCCAAACAAAGATCTGACAGAT 552  
 1139 yrttyr 1140  
 553 ATTAC 557  
 seq\_name: gb\_est89:BF528156

seq\_documentation\_block: 725 bp mRNA  
 LOCUS BF528156  
 DEFINITION 602042746F1 NCI\_CGAP\_Brn67 Homo sapiens cDNA clone IMAGE:4180378  
 5', mRNA sequence.

ACCESSION BF528156 GI:11615519

VERSION BF528156

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 725)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: David N. Louis, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.









CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
STREET: 1601 Market Street Suite 720  
CITY: Philadelphia  
STATE: PA

COUNTRY: USA

ZIP: 19103-2307

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/16216

FILING DATE:

CLASSIFICATION:

Prior APPLICATION DATA:

APPLICATION NUMBER: US 08/353,700

FILING DATE: 09-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36,252

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 10136 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

PCT-US95-16216-2

alignment scores:

Quality: 439.50 Length: 1256

Ratio: 0.668 Gaps: 49

Percent Similarity: 52.389 Percent Identity: 19.904

alignment block:

US-09-538-396-2 x PCT-US95-16216-2 ..

Align seg 1/1 to: PCT-US95-16216-2 from: 1 to: 10136

51 CysThrGluGluLeuProPheSerSerSerGlyHisThrPheValH 67

5619 TGCCTGGGAGACAGTCCCGCATACCAAT.....TA 5650

67 SASPTOLysValAlaGlyGluThrGlyGlyGlnIleLeuLeu 84

5651 TGAGCGCT.....CCAGGGAGAAATAAACCCAGGCGCTTCAGAA... 5690

84 rghpelysThrAlaIleGlyAspValValCysIleArg.....Ser 98

5691 .....TGCATTTCTGAATTGCA 5708

99 PheGlnLeuThrGlnLysAlaSerLysMetGluPhe.....Lys 111

5709 TTTCTGGTCTTAATGCTTTGTTACATGATTTCCCGGGAGATCAGA 5758

111 AlaIleGluSerValLeuGlnThrIleAsnProIleThrGlyGluVal 128

5759 AGATATCCATATCTTCMAATCGCGGTAAAGACATCAATGAAATTT 5808

128 ALGSLseuSerThrArgCysAlaAspMetAspArgGluIleProAlaLeu 144

5809 TGAGATTACTTCATGATAGAGACCGTGACAGAAAAGTTGAAAGTTTG 5858

145 MetGlyVal.....SerLysAlaValLeuGluAsnValI 156

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

5859 CTAATGAATGAAGAATAGACTCAAACTCCATTTAGAGAGTACA 5908

156 ephValHisGlnAspLysSerAsnThrProLeuGlnAsp...Proser 172

5909 ACTATGACCAAAATTAAGACATGATGAAATTAAGTGGG 5958

172 hLeuLysLysLysPheAspAspIlePheSerAlaThrArgThrLys 188

5959 AACTTAAGAAAAGAACTCAGATTTA.....AGTAA 5990

189 AlaLeuGluValIleLysLysLysHisLysAspGlnMetGlnIleLys 205

5991 AAATGGAAATATTTCTGATGATCAGAGAGTACTCCAGAGTACA 6040

205 sThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspGlnLan 222

6041 AACTTGAAGGCGCTCAATCTGATTTAGAAATGATGATGATATCAT 6090

222 LysLysLeuArgGluAsnIleAlaGlnAspGlnGlySerAspAlaSer 238

6091 CA.....CGTGAAGATATTGAGATATATGCGCAAGGTGAATGACAGC 6134

239 .....LysSerGlnMetGlnLysLysG 247

6135 TGAAGAGAGATTTCTGATGATGAGAAATGAGCTGATGATGATGATC 6184

247 uLysIleCysGlyThrGluArgGluIleLeuGlnMetGlnThrSerLeu 264

6185 GGAGAAAGCTAGCATTTGATGATGAGAGCCCTTACTGAGGCTACTTGG 6234

264 sPGLLeuArgArg..... 268

6235 AGTAGTTCAACAGAGAGATGTTAGAAAAGCAATGAATAATAG 6284

269 .....LeuGlnGlyGlnIleAspIleLysAlaThrG 279

6285 CAGAGGTATTGCTGCTGATGAGAAAGACTCTCAGTGTCAAGAGTGA 6334

279 uArgSerThrLeu.....LeuThrGlnGlnHisGul 290

6335 GAGAAACAGCTCTGAGAGATTAATGATCTCAAAAACCCACCG 6384

290 LysLeuAlaIleLeuSerGlnGluAsnGluAspThrAspGlnGluMet 306

6385 CACTGATCACTGCTGTAATAAATGAAGGAGAAACACAGACTTGG 6434

307 GluThrGlnThrLysPheGluGluArgIleAlaLeuLeuGlnThrLys 323

6435 TCTCATCAAACTGAGTGTCTCATTCATGAGTGGCAGAGGAGAGT 6484

323 eSerLysLeuValArgAspMetAspAspGluAlaSerThrSerSerVal 340

6485 GAGAGAAAGAGAGAGACTCTTCAGACTTGTCTGATGTGAGTGA 6534

340 euseLysGlnAsnSerGlnLeuThrHisGlnIleGlyLysLeuGlnAla 356

6535 TGTTAAAGACAAATCATCTCCAGGAAAGCTGCAGAGTTTGAAGA 6584

357 GluAlaAspAlaHisLeuThrMetLysHisGluArgAspSerAspIle 373

6585 GACTCAGAGGAGACTGTCTTGACAAATGATGAGCTGGAACCAATATGC 6634

373 sAsnIleCysThrLysHisAsnLeuGlyProValProGlnHisProPhe 390

6635 ACAACTGATTAAGAG..... 6650

390 hAsnAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSer 406

6651 ..AAGATTTCTGCTCAGAGATCTGAAGCTGCAGAGCCAGACTGAT 6698

407 SerLeuGluAsnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLe 423

6699 GAATCAGATTATGAAGAGCTGAATGTCTCAAGGCGCTGAGGCGGCACT 6748

723 uArSerrLyAlaLeuLalMetGluSerSerAsnAlaGluAlaLeuPhcG 740  
||||| :||| ||||| :|||  
7376 GAGGGCCCGCTTGAACTCATATAAAGAAGCAGCTC...TGTGCTTCAC 7422  
740 InGlnLeuApRyLSerLysArg.....ThrIleTyraSPalATyVal 753  
||||| :||| :||| :||| :|||  
7423 AACACACTGAAGCAAGAGCACATCATGCCAGATTCTTAAGGTTAGACG 7477  
754 LysLeuValJuglUthrIleProLeuAlaGluLysAspLeuAsnInH 770  
:::||||| ::| :||| :||| :||| :|||  
7473 GAGAACCTTGAAAGAGAGCTTAGAGATTAGCCAGACAAC...CAAGAGA 7511  
770 sLeuAlaAspGluSerGlnLysAlaGlnAlaPheAspLeuGlyY 787  
||| ||| ||||| :||| :||| :||| :|||  
7520 TGCAGCTCTTGAGCAGAGAAATTCCAA..... 7547  
787 allLeuAlaHisValGlnMetLaspArgAspAlaValGluAlaLeuGln 803  
:||||| :||| :||| :||| :|||  
7548 ....GGAGAGCTAGAGACCCTTAAGCAAAAATAGAAAGGATGACCCA 7592  
804 ProThrasPrTHleasPrghisVal..... 812  
7593 AGCTGTGAGAGCTGTGGAATTAGATTGTTACTATPAAGTCAGAAAAA 7644  
813 .....HisGluIleGlnGlnLeuValLysGluValGluAspLeuG 826  
7643 AAATGTGACAAAGAAATTCACAAAAGCAAGCAAGCAATATCTGAATTAG 7692  
826 lutyralaleuAspSerSerGlyArgGlyVal.....LysSerLeu 839  
||| :||||| :||| :||| :|||  
7693 AA..ATAATAATTATATCATTTTGAAAATATTTTGCAAGAAAAGACA 7738  
840 GluGluIleGlnLeuLeuAsnPheLeuGlnArgThrArgAspThrLe 856  
7740 GAGAAAGTCACAGATG.....AAAGAAAATCAAGAC 7771  
856 ulileValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGlu 873  
7772 TGCCATGAGGATG....CTTCAACACAAATAAAGAGCTCATGTAG 7815  
873 spketerSerAlaGluValArgTPHLeAsnAlaArgGluGlu...Lys 888  
7816 GAGTGCACAGCTCG.....CATATGACCAAGAAAGCTGTAG 7853  
889 ValLysAlaSerSerIleLeuGluArgPheGlnLysSerGluGluLe 905  
7854 GCCAAAGACAGAAAT..... 7868  
905 uValleuLeuAlaGluGluLysGluGlnLeuIleValGluLysLysLeu 922  
7869 .....CTTAGTAGCAAGTAGAGTGTCTTGAACTGAGAAAGCTCAGT 7911  
922 euclugluSerLeuAspPro.....LeuSerLys 931  
7912 TGCTACAAAGGCTTGATGAGGCCAAAATTAATTATATTCTTTSCAAATCT 7961  
932 gLuLysGluSerLeuLeuGlnGlnLutyraAsnAlaLeuLysGlnLysLeuAs 948  
7962 TCAGTGAAGGCTCATTCAGAAAGATAGAAAGTGCACAGCAAGAACTGGA 8011  
948 pGluGluLutyriSGlnLeuAlaGluArgLysArgGluPheGlnGlnGlu 965  
8012 GAAGAGAGTGAAGAAACACAGTAGACTAAAAATCAATTCAAGACCAAG 8061  
965 euAspAlaLeuGluArgLysAsnMetLysIleLysGlyTYTyLeuAspSer 981  
8062 ACACAGCTTCTC..... 8072  
962 LysLysAsnGluLysLeuLysGluLeuGlnLutyraHisValLeuCysHI 998  
8073 .....TTAAACTGTGCCAGGTGGAAGAGAGCAACCAACTTGGAA 8113  
998 sSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgIleSerAlaGlu 1015

```

8114 GGAGCA.....AACTTAGAAGTGAAGAACTGACGCTGAAAT 8151
      |||
1015 euasnlyserlysglyleuenglnsglynglnleuylsarghpn 1031
      |||
8152 TGGAGCAGAAAGTCAAGTCTCAATCAAA.....AATGCCCT 8192
      |||
1032 ilasparapnleuys.....tyrargylsthrlysal 1043
      |||
8193 TTGCGAGCAGACATTAGAGTCTCAGAGTCTTCAAGAACTTAGAGAA 8242
      |||
1043 aaspyalgluInleuthrargaspillegluserleuInuargleuL 1060
      |||
8243 TGAGCTTAATGGACAAAATGGACAAAATGCTCTTTGTAAGAAA... 8288
      |||
1060 euserileglyserleuserialleglualasrleuylsarghissr 1076
      |||
8289 .....GTAAACAAAATGACTGCAAAAGAACTGAGCTCAGAGGAAATG 8333
      |||
1077 glnglu.....lysgluarglyleuanserGI 1085
      |||
8334 CATGAGATGGCAGACAAAACAGAGAGTGCAGAGAACTCAGTGAGA 8383
      |||
1085 uphasnargtrprnglnlythrleuserValtycginseranlisertL 1102
      |||
8384 GAAAATATAGCTAGCTGAGAGAGTTCAGTCTGTGGAGAAATAAAGA 8433
      |||
1102 yshlelysglnleuylsleuserglnlytyrlyaspilleglulysarg 1118
      |||
8434 CGAGCAAGATCAATGTAAG..... 8453
      |||
1119 tythransglnphelenglnleuylsthrngluMetAlaslnlysas 1135
      |||
8454 .....GACCTCACACTAGAAAATAGTGA..... 8477
      |||
1135 pleuspargtyrtythrAlaleuAsplysAlaleuMetArpPheHLS 1152
      |||
8478 .....TTGAGAGAGAGCTTACATTGCATGCACACA 8505
      |||
1152 ermetlyswetgluInleuInlysllelelysglnleuInleuIn 1168
      |||
8506 AAGACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8555
      |||
1169 thrtyrarglyglinsrpleasp.....tyrileserileasnsr 1183
      |||
8556 CAGCTACGCTTCATGAGAGCTGAAAAGAACCCAGAGCTTTGCTTTGGA 8605
      |||
1183 psergluglyalaglythrargserlyserlyrargValValmetgIn 1200
      |||
8606 CACA.....AACAAACAGTATGAGAGAGAGAGAGAGAGAGAG 8637
      |||
1200 hrglyAspAlaleuInleuMetarglyargCysSerAlaInlyIn 1216
      |||
8638 CATACCGAGAGAAATGACTCTTAAGAAGAGATGCTCAGTCCACAGAG 8687
      |||
1217 valleuAlaserleuile 1222
      |||
8688 CTGAGAGATGAGCTTTTA 8705

```

seq\_name: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:US-08-139-937-12

seq\_documentation\_block:

Sequence 12, Application US/08139937

Patent No. 5821070

GENERAL INFORMATION:

APPLICANT: LEE, MEN-HWA

APPLICANT: SHAN, BEI

TITLE OF INVENTION: CELLULAR GENES ENCODING

TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:

ADDRESSEE: CAMPBELL AND FLORES

STREET: 4370 LA JOLLA VILLAGE DRIVE

```

? CITY: SAN DIEGO
? STATE: CALIFORNIA
? COUNTRY: USA
? ZIP: 92122
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/139,937
? FILING DATE: 20-OCT-1993
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/979,156
? FILING DATE: 20-NOV-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: CAMPBELL, CATHERYN
? REGISTRATION NUMBER: 31,815
? REFERENCE/DOCKET NUMBER: P-CJ 9370
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 619-535-9001
? TELEFAX: 619-535-8949
? INFORMATION FOR SEQ ID NO: 12:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4868 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-08-139-937-12

```

```

alignment_scores:
  Quality: 419.50      Length: 1194
  Ratio: 0.655        Gaps: 46
  Percent Similarity: 53.601  Percent Identity: 20.184

```

alignment\_block:

US-09-538-396-2 x US-08-139-937-12

Align seg 1/1 to: US-08-139-937-12 from: 1 to: 4868

```

113 ilegluserValleuInleuInthrleasnProHlsThrGlyInlyValcy 129
      |||
51 ATCCATATCTTCAACTCGGGTAAAGACATCAATGAGATTTAG 100
      |||
129 sleuserlyrargCysAlaaspMetaspArgluileProAlaleuMetG 146
      |||
101 ATTACTTCATGTGATAGAGAGACCGGACAGAGAAAGTGAAGTTGCTTAA 150
      |||
146 lyval.....SerlyslalValleuInleuInVallelPhe 157
      |||
151 ATGAATGAAGAAATTAAGCTCAAAATCCATTTACAGAGGTTAACTA 200
      |||
158 valHISGlnaspGlnSerAsnTrpProleuGlnasp...ProSerThrle 173
      |||
201 ATGACCAAAATTTGAAAGCATGCTAGAAATTTGAAAAAATAGTTGGGGAG 250
      |||
173 ulslyslusPheaspPilePheSerAlaThrargtyrThrlyslalL 190
      |||
251 TAGAAGAGAAAGCTCAGATTTA.....ACTGAAAAAT 282
      |||
190 eugluVallelyslusleuInlyslasrPglMetGlnleuInlysthr 206
      |||
283 TCGAATATTTTCTGTGATCACAGAGATTACTCCAGAGAGTAAAGAACT 332
      |||
207 PheArgleuLyLeuGlnaspInleuInthrVallyAspGlnAlaHlsly 223
      |||
333 TGTGAAGCCTCAATTCATTTAGAAATGCTGCAGATNAATATCA... 380
      |||
223 sleuArgluAsnleAlaGlnaspInleuInlysserAspAlaser... 238
      |||

```



```

791 aGlnMetAspArgSP..... 796
2008 TTTTCGCAAGAAAAAGAGCAAGAAAGATGACATGCAAAAGAAATCAAGC 2057
797 ..AlaValGluAlaLeuLeuGlnProThrAspThrIleAspArgHisVa 812
2058 ACTGCCATGAGATGCTTCAAAACAAATTAAAGAGCTCAATGAGAGAGT 2107
812 LHisGluLeu.....GlnGln 818
2108 GGCACCCCTGCAATATGACCAAGAGCCCTGTAAGCCAAAGAGAGAGATC 2157
818 euVallyGluValGluAspLeuGlnTyrAlaLeuAspSerSerGlyArg 834
2158 TTAGTAGTCATAGTAGTGTCTTGAACTT.....GAGAGGCTTCAG 2198
835 GlyVallySerLeuGlnGlnLeuLeuGlnLeuAsnPheLeuGlnArg 851
2199 TTGCTACAAAGCCCTGATGAGGCCCAAAATATATATTGTTTTCATATC 2248
851 GThrArgAspThrLeuIleValGluValAspAspLeuArgAspGlnHis 868
2249 TTCAGTAGATGCTCATTCAGAGAGTAGAGTAGAGAGAGAGAGAGAGAG 2298
868 IGMetLeuAsnGluAspMetSerSerAlaGlnValArgTrpHisAsnAla 884
2299 ACAAGAGAGATGAGAAATCACTAGATGAAATCAATTCATCAAGACCA 2348
885 ArgGlnGluVallyValAspSerSerIleLeuGlnArgPheGlnLys 901
2349 GAGCAGAGCTGTCTTAACCTGCTCCAGAGGAGAGAGAGAGAGAGAGAG 2398
901 rGlnGlnGlu.....LeuValLeuLeuAlaGlnGlu 912
2399 GAGAGAGCAAACTTAGAAGTGAATCTGACAGTGAATGAGAGAGAG 2448
912 YGlnGlnGlnLeuIleValGlnLys..... 919
2449 ACATCCATGCTGACATCCAAATGCTCTTTCAGAGACATATAGAA 2498
920 .....LysLeuLeuGlnGlnGlnSerLeuAspProLeu 930
2499 GTCGTGACAGATTCCTTACAAAGATCTAGAGATGAGCTTGAATGACAA 2548
930 rlyGlnGlnGlnSerLeuLeuGlnGlnGlnLysAlaLeuLysGlnLys 946
2549 AATGACAAATGCTCTTGTGAAAGTAAACAAATGACTGCAGAAAG 2598
947 .....LeuAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 960
2599 AACTGAGCTGAGAGAGAGAGATGATGATGACACAGAAACACAGAG 2648
961 PheGlnGlnGlnLeuAspAla.....LeuGlyArgLeuAs 972
2649 CTGCAAGAGAGACTCAGTGAAGAGAAATAGCTAGCTGAGAGAGTGA 2698
972 nMetLysIleLysGlyTyrLeuAspSerLysAsnGlnLysLeuLysG 989
2699 GTTACTGTTGGAAGAAATTAAGAGAGCAAA.....GATCAATGGAAG 2742
989 lueGlnGlnLysValLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 1002
2743 AGCTC.....ACACTAGAAATAGTGAATTTGAAGAGAGCCTA 2780
1003 SerCyMetAlaLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1019
2781 GATGATGATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2827
1019 rGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1031
2828 GGAATAGCTGATATACAGTACGCTTATGATGAGCTGAAAGAGAGAG 2877
1032 .....IleAspAsnLeuLysTyrArgLysThrLysAlaAsp 1044

```

```

2878 AGCTTTCCTTTGGACACAAACAAACAGTAT.....GAAGTAGAA 2918
1045 ValGlnGlnLeuThrArgAspIleGlnSerLeuGlnGlnGlnGlnGln 1061
2919 ATCCAGATATCCGAGAGAAATGACTTCTTAAGAGAGATGCTC..... 2963
1061 rIleGlySerLeuSerAlaIleGlnAlaAspLeuLysArgHisSerGln 1078
2964 .....AGTCACAGAGAGCTGAGATGAGCTTTTAAGTCTAGT.... 3002
1078 luyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1094
3003 ..AAGAGAGCTCAAT.....AATCATGCAAGCTACTACTAG 3041
1095 ValTyrGlnSerAsnIleSerLysHisGlnGlnGlnGlnGlnGlnGln 1111
3042 ATTTTG.....GAGAAATTGAAGAGAGAGAGAGAGAGAGAGAG 3067
1111 nTyrLysAspIleGlnLysArgTyrThrAsnGlnPhe..... 1123
3068 GATGACATCTA.....AATATGTAATCAGTTGAAGAGAGAGAGAG 3111
1124 .....LeuGlnLysThr..... 1128
3112 AACGTCCCGAGGAGAAATGAAGTTGTTGATCAATCTGTAACAGCTG 3161
1129 .....ThrGlnMetAlaAsnLysAspLeuAspArgTyrTyrThrAl 1142
3162 GAGAGAGAGAGAGAGATCTCAGAGAGAGAGAGAGAGAGAGAGAGAG 3211
1142 AleuAspLys.....AlaLeuMetArgPheHisSerMetLys 1155
3212 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3252
1155 etGlnGlnLysAsnLysIleIleLysGlnLeu 1165
3253 TCGATGAATTAACACTGAGATCAAGAGAGAGAGAGAGAGAGAGAG 3284

seq_name: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq.PCT-US93-11310-12

seq_documentation_block:
: Sequence 12, Application PC/TUS9311310
: GENERAL INFORMATION:
: APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
: TITLE OF INVENTION: CELLULAR GENES ENCODING
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CAMPBELL AND FLORES
: STREET: 4370 LA JOLLA VILLAGE DRIVE
: CITY: SAN DIEGO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/11310
: FILING DATE: 19-NOV-1993
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: CAMPBELL, CATHRYN
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: PP-CJ 9790
: TELEPHONE: 619-535-9001
: TELEFAX: 619-535-8949
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:

```

LENGTH: 4868 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 PCT-US93-11310-12

alignment scores:  
 Quality: 419.50 Length: 1194  
 Ratio: 0.655 Gaps: 46  
 Percent Similarity: 53.601 Percent Identity: 20.184

alignment block:  
 US-09-538-396-2 x PCT-US93-11310-12

Align seg 1/1 to: PCT-US93-11310-12 from: 1 to: 4868

113 ILeGUsErVAlLeuGInThrIleAsnProHISThrGlyLysValCy 129  
 51 ATCCATAATCTTCAACCTCGGCTAAAGACATCAAAATGAGAAATTGAG 100  
 129 sLeUsErTyRarGysAlaAspMetAspArgLuiLeProAlaLeuMetG 146  
 101 ATTACTCATGTGATAGAGACCGCTGACAGAAAGTTGAAAGTTGCTTA 150  
 146 lyVal.....SerLysAlaValLeuGluAsnValIlephe 157  
 151 ATGAATGAAGAAGATTAGACTCAAAACCTCATTTACAGAGAGTACAACTA 200  
 158 ValHISGlnAspGluSerAsnTrpProLeuGlnAsp..ProSerThrIe 173  
 201 ATGACCAAAATGTAAGACATGCATAGAAATGAAAAATAGTTGGGGAAC 250  
 173 uLySLysLysPheAspAspIlePheSerAlaThrArgTrpThrLysAlaL 190  
 251 TAAGAAAGAAACCTCAGATT.....AGTGAATAAT 282  
 190 euGluValIleLysLysLeuHisLysAspGlnMetGlnLuiLeLysThr 206  
 283 TGGAAATATTTTCTGTGTGATCACAGAGACTTACTCCAGAGAGTAGAAACT 332  
 207 PheArgLeuLysLeuGlnAsnLeuGlnThrValLysAspGlnAlaHisLys 223  
 333 TCTGAGAGCCCTCAATCTGATTAGAAATGATGCAGATAAATCAACA... 380  
 223 sLeuArgGluAsnIleAlaGlnAspGlnLuiLysSerAspAlaSer.... 238  
 381 ...CGTGAAGATATTGGAGATAATGTGGCCAGAGTGAATGACAGCTGGA 426  
 239 .....LysSerGlnMetGlnLuiLeuLysGluLys 248  
 427 AGGAGAGATTTCTGTGATGGAATAATGAGCTGAGTAGATCAGATCGAGAG 476  
 249 IleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspIle 265  
 477 AAGCTAGCATGAGCATGAAGCCCTCTACCTGAGAGCTGACTTAGAGGT 526  
 265 uLeuArgArg..... 268  
 527 AGTTCAAAACAGAAAGCTATGTTAGAAAACACATGAATAATACACAGA 576  
 269 .....LeuGlnGlyGlnIleAspIleLysAlaThrGluArg 280  
 577 AGGTATTCTGCTGCTGGAAGAAGACTCTCAGTGTCTCAAACTGAGAGA 626  
 281 SerThrLeu.....LeuThrGlnGlnHisGluLysLe 291  
 627 AACCAAGCTTCGTGAGAGATTAGATCTATCTCAAAAAAACCAAGCCGAC 676  
 291 uAlaIleLeuSerGluGluAsnGluAspThrAspGluLuiLeuMetGluTr 308  
 677 GGATCAGTTGTCTGAAAAAATGAGAGAAACACAGAGCTTGAGTCTC 726

308 rPgInThrLysPheGluGluArgIleAlaLeuLeuGlnThrLysIleSer 324  
 727 ATCAAGAGAGAGTGTCTGCATTCATTCAGGTGCGAGAGCAGAGGTGAG 776  
 325 LysLeuValArgAspMetAspAspIleAlaSerTrpSerSerValLeuSe 341  
 777 GAAAGACGGAACCTCTCAGACTTGTGCTCTGATGAGAGACTCTT 826  
 341 rlyGlnAsnSerGluLeuThrHisGlnLuiLeuLysLeuGlnAlaGlu 358  
 827 AAAAGCAAACTCATCTCCAGAAAGCTGCAGAGTTTGGAAAAAGACT 876  
 358 lAspAlaHisLeuThrMetLysHisGluArgAspSerAspIleLysAsn 374  
 877 CACAGGCACTGTCTTGCACAAATGTGACCTGGAACCAATTCACACA 926  
 375 IleCysTrpLysHisAsnLeuGlyProValProGlnHisProPheThrAs 391  
 927 CTGAATAAAGAG.....AA 940  
 391 nAspValAlaMetAsnLeuThrAsnArgLileLysAlaArgLeuSerTrL 408  
 941 AGAATTGCTTGTCAAGGAATCTGAAGCCTGACGCGCAGACTGAGT 990  
 408 euGlnAsnAspLeuAspLysLysLysSerAsnGluAspGlnLeuAsp 424  
 991 CAGATTATGAAGCTGAATGTCTCCAGAGCCTTGAGAGGCCCATGTG 1040  
 425 ValLeuTrpLysHisTrpLeuLysIleAsnAlaArgTrpSerGluValAs 441  
 1041 GAGAAAGCGAGTGTCCATTCAGAGCTGACCTCAACAGAGAGAGTGA 1090  
 441 pElGlnIleGlnSerLysIleGluSerMetSer..... 452  
 1091 T...CAGCTGAGAGAGCGATCGAGAACTGAGAGTTCGATGAGCGCG 1137  
 453 .....GlyIleLeuArgArgArgLysAspLysGluLys 463  
 1138 AGCAAAAGAGCAGCTGCACATCGCAGAGAACTGAAAGACCGAGCGG 1187  
 464 GluArgAspAlaIleAlaGluLeuSerLysPheAsnLeuSerArgIle 480  
 1188 GAGAAATGATTCCTTAAGTA.....AAAGT 1213  
 480 eaSpGluArgGluArgHisMetGlnIleGluValGluArgLysThrLeuA 497  
 1214 TGAAGACCTTGAAAGGGAATGCGAGATGTCAGAAAGAAACCGAGACTAG 1263  
 497 la...LeuGluArgAspTrpAspSerIleIleSerGlnLysArgThr 512  
 1264 TGAATCTGATGCCGAGAAATCCAAAGCCAGAACTAGAGACTTAATAACA 1313  
 513 GluValTrpSerLeuGlnLuiLysIleLysValLeuLeuArgGluLysAs 529  
 1314 CAATAGAGAGATGGCCAGAAACCTGAAAGTTT....GAATAGA 1357  
 529 rIleLeuAsnArgAsnAlaAspIleArgValLysLysLeuLysLysAs 546  
 1358 CTTGTC..... 1364  
 546 sPAlaLeuGluSerLysAspLysLeuAsnGluLuiLeValAsnGlnHis 562  
 1365 ..ACGTTAAGCTGAAAAAGAACTGCACAAACAAATACAGAAAAA 1412  
 563 LysAspLysIleLysLysValLeuArgLysArgAsnProPheGluLysAs 579  
 1413 CAAGCTCAGTGTGACAGACACAGCAAGTACTCTTCAATTTAAAGTCT 1462  
 579 pMetLysLysGluIleAsnGlnAlaPheTrpProValAspLysGluTrpA 596  
 1463 GTTAGAA.....GAAAGGAGACAAAG 1482



3112 AACGTGCCAGGGAATAAGTGTGTGATCAATCTGTAACAGCTG 3161  
1129 .....ThglumetalasnllyspaleuspargtyrtyrthAl 1142  
1142 .....  
3162 GAAGAGCAAAAGAGATGACTGACAGAAAGACTCTCTCAACTTCAAGCTGC 3211  
1142 aleuasplys.....Alaleumetarphhehlsersmellysm 1155  
3212 ACAGAGAGAGCAAAACAGCTACTGTTATG.....GATACCAAG 3252  
1155 etglugluileasnllysllellelysgluLeu 1165  
3253 TCGATGATTAACAACGTGATCAAGAACTG 3284

seq\_name: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:US-08-687-080-54

seq\_documentation\_block:

Sequence 54, Application US/08687080

Patent No. 5965427

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

NUMBER OF SEQUENCES: 175

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,080

FILING DATE: 17-JUL-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/592,126

FILING DATE: 26-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111.30

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 543 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: RAD50-3

US-08-687-080-54

alignment\_scores:

Quality: 391.50

Ratio: 2.663

Percent Similarity: 79.032

Percent Identity: 43.548

alignment\_block:

US-09-538-396-2 x US-08-687-080-54

Align seg 1/1 to: US-08-687-080-54 from: 1 to: 543

71 ValAlaGlyGluThrGluThrLysGlyGluIleLysLeuAsgPheLysTh 87  
1 GTTGCCTCAAGAAACAGATGTGAGACCCAGATTCGTCGCAATTCGCGA 50  
87 TAlaAlaGlyLysAspValValCysIleArgSerPheGlnLeuThrGln 104  
51 TGTCAATGAGCAACTTATAGCTGCAAAAGATCATCTGCTGCTACTCAGA 100  
104 yAlaSerLysMetGluPheLysAlaIleGluSerValLeuGlnThrIle 120  
101 AAAGCAAAAGACAGAAATTTAAACTCTGGAAAGAGTCAAT...ACTTGA 147  
121 AsnProHisThrGlyGluLysValCysLeuSerTyArgCysAlaAspMe 137  
148 ACAAAAGCAT...GTGCAAAAGCTCAGTCTGAGCTCTAAGTGCAGAAAT 194  
137 TAspArgGluIleProAlaLeuMetGlyValSerLysAlaValLeuGlu 154  
195 TGACCGAGAAATGATCATCTCTGGGGTTCCAAAGCTGCTGAATA 244  
154 srValIlePheValHisGlnAspGluSerAsnTrpProLeuGlnAspPro 170  
245 ATGTCAATTTCTGTCATCAAGAAATTCATTAATGGCTTTAAGTAAGCA 294  
171 SerThrLeuLysLysPheAspIlePheSerAlaThrArgTyTh 187  
295 AAGCCTTGAGCAAAAGATTTGATGAGATTTTTCAGCAACAAGTACT 344  
187 rLysAlaLeuGluValIleLysLysLeuHisLysAspGlnMetGlnI 204  
345 TAAAGCCTTAGAAACACTCGGAGAGTACGACACAAAGCTGAAGA 394  
204 lAlThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspGln 220  
395 TAAAGAAATATCAATGAGACTAAATATCTAACCAATATTAAGAAAA 444  
221 AlaHisLysLeuArgGluAsnIleAlaGlnAspGlnLysSerAspAl 237  
445 GCTTGAGATTCGTGATCATCAATTA.....AG 473  
237 aserLysSerGlnMetGluGlnLeuLysGluLysIleCysGlyThrGlu 254  
474 TAAGAGCCCAAGTTACATCTTCAAGAAATTTGCAAAATCTTATGAA 523  
254 rGluIle 256  
524 ATGAAGCTT 531

seq\_name: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:US-08-466-390-3

seq\_documentation\_block:

Sequence 3, Application US/08466390

Patent No. 5686562

GENERAL INFORMATION:

APPLICANT: TOKRATLY, GARY

APPLICANT: LIDGARD, GRAHAM P

TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE

INTERIOR NUCLEAR MATRIX

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: 125 HIGH STREET

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,390



FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER BSO, EDWARD R.  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: MTP-013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6306 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..6306  
PUBLICATION INFORMATION:  
AUTHORS: COMPTON, DUANE A  
AUTHORS: SZILAK, ILYA  
AUTHORS: CLEVELAND, DON W  
TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR  
TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR  
TITLE: SEGREGATION OF PROTEINS AT MITOSIS  
JOURNAL: J. Cell Biol.  
VOLUME: 116  
PAGES: 1395-1408  
DATE: 1992  
US-08-466-390-3

alignment scores:  
Quality: 376.50 Length: 1486  
Ratio: 0.510 Gaps: 62  
Percent Similarity: 49.664 Percent Identity: 19.314

alignment block:  
US-09-538-396-2 x US-08-466-390-3

Align seg 1/1 to: US-08-466-390-3 from: 1 to: 6306

```
103 GlnYsAlaSerLysMetGluPheLysAlaIleGluSerValLeuGln.. 118
102 CAGGGGCCAGCCACCTGAGCCAGAGAGCTGAGAGCTGCGTGACAA 851
119 ..ThiLeaSPROHisThrGlyGluLysValCys... 129
852 GAATAGAGACCTTACCATCGCGCTCATGAACCTGAAGCAGTGCACAG 901
130 ..LeuSerTyrArgCysAlaAspMetAspArgGluIleProAlaLeuMet 145
902 ACCTGAGAGAGAGAGAGAGAGAGATGATCGCAATATC..... 939
146 GlyValSerLysAlaValLeuGluAsnValIlePheValHisGlnAspG 162
940 .....AACCAAGCTTCGAGGAGAGATGAGAA 965
162 uSerAsnTrpProLeuGlnAspProSerThrLeuLysLysLysPheAspA 179
966 CTTTCCTTAAAGCTGCGGAGATTGCCAGTCACTGCAGCAGCAGTACAG 1015
179 sPLePheSer.....AlaThrArgTyrThrLysAla..... 189
1016 ATGCCCTCAATGAGCTGAGCGAGAGAGACAGCAAGCCACTCAGAGAGTGG 1065
190 LeuGluValIleLysLysLeuHisLysAsp.....GlnMetGlnG 203
1066 CTAGAGAAAGCAGGCCAGCTGGAGAGAGAGCTCAGCAGCAGCCCTCAGGA 1115
203 uLLeysThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspG 220
```

```
1116 CAGAAATGCTTGAAGAGAAAGAAATCTTCAG.....GGA 1156
220 InAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnGluLysSerAsp 236
1157 AACCTTCACAGCTGAGAGAACACTTGTCCAGCTGCAGAGATACCAACCACC 1206
237 AlaSerLysSer.....GlnMetGlnGlnLeuLys 246
1207 CAGAGAAAGGCGAGAGTCTGGTGTATGTCTTGCAGCTGGAACCTTGA 1256
246 sGluLysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerL 263
1257 GCAAGAGGCGACCACTTGTCTGCAAAACACACAGCTCCAGCCAGCAGG 1306
263 euAspGluLeuArgArgLeuGlnGlnIleAspIleLys..AlaThr 278
1307 TAGAGATGCTGAGAGCTGAGGAGGAGCCAGAGAAAGCCAGCTGCTTCT 1356
279 GluArg.....SerThrLeuLeuTh 285
1357 GAGCGGGCCACTTCGAGAGAAAGAGAGAGAGCTGTACTACCTGATCAG 1406
285 rGlnGlnHisGluLysLeuAlaIleLeuSerGluGlnAsnGluAspThrA 302
1407 TGACCTGAGAGCTCCATCTCAACCTCAGCCAGCCAGAGAGAGCTGG 1456
302 sPGLuLLeuMetGluTrpGlnThrLysPheGluGluArgIleAlaLeu 318
1457 AGCAGGCTCCAGGCTCATGGGGCCGTTGACTGCCAGGTGCTCT 1506
319 LeuGluThrLysIleSerLysLysValAlaArgAspMetAspGluAlaSe 335
1507 CTGACCTGAGCTACCACTCAATCCACCATCCAGCAAGAGAGATCA 1556
335 rTyrSerSerValLeuSerLysGlnAsnSerGluLeuThrHisGluIleG 352
1557 AGAAGCTGCTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606
352 LysLys..LeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHis... 366
1607 CACAGACCTCCACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
367 .....GluArgAspSerAspIleLysAs 374
1657 GTGAGCAGAGCTAAGCAGTACGCTGAAGAGAGAGAGAGAGAGAGAGAG 1706
374 nLecTyrThrLysHisAsnLeuGlyProValProGluHisProPheThrA 391
1707 GTTACGCGAGAGAG.....CAGAGAGCACTAGAGC 1735
391 snAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSerSer 407
1736 AGGACCATGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTTC 1785
408 LeuGluAsn.....AspLeuLeuAspLysLys 416
1786 TTAAAGAGAGCGGAGATCGGCTCTCAAGCAGCTGAGAGAGAGAGAGAG 1835
416 sLysSerAsnGluAspGlnLeuAspValLeuTrpLysHisTyrLeuLysI 433
1836 GAAGGCT.....GCCAAGCTGAGAGATTCGAGCAGCAACTTCAGTGG 1879
433 LeuAsn.....AlaArgTyrSerGluValAspGlyGln... 443
1880 CTAAATGAAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1929
444 .....IleGlnSerLysIleGlnSerMetSerGlyIleLeuArg 456
1930 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1979
456 g.....ArgArgLysAspLysGluLysGluArgAspAlaAlaGluValG 471
1980 GACAGCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029
```

704 roAspEugluAspLuPhelVallyLysLysGlnArg...MetGlnAAsnSer 719  
 |||  
 |||  
 2834 GAGCAAGACAGCCGCCAGGGTGGAGAAAGCAACAGGCGCCCACTTCTCC 2883  
 |||  
 |||  
 720 SerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSerAsnAlaG1 736  
 |||  
 |||  
 2884 AGCAACACAGGCGCCCTCAGGCTATGGAGGGGAG...GCAGA 2924  
 |||  
 |||  
 736 uAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyrAspAlaTyrV 753  
 |||  
 |||  
 2925 GCAGATGGGCATATGAGCTGCAACGCGCTCGG...G 2956  
 |||  
 |||  
 753 aLlyLeuValGluGluThrIleProLeuAlaGluLysAsnLeuAsnGln 769  
 |||  
 |||  
 2957 CCGCCTGATGGAGAGCCAGGGCGCAGCAGAGAGAGACCTGGGCACTCG 3006  
 |||  
 |||  
 770 .....HisLeuAlaAspLuSerGlnLysAlaGlnAla.. 780  
 |||  
 |||  
 3007 GAAGAAGAGGTGGCGCGCTGACCCAGAGAGGGGGCGCTGCCACGCTGA 3056  
 |||  
 |||  
 780 ..... 780  
 |||  
 |||  
 3057 CGTGGCCCTGGAGAAAGGGCGCCAGACAGAGCTTGAGATGGCGCTGCAGA 3106  
 |||  
 |||  
 781 .....PheAspAspLeuGluValLeu 788  
 |||  
 |||  
 3107 AGCCCTCAACAGCAGCAGCGCTGGAGTTCGCTACCTCCACAGGCACTGG 3156  
 |||  
 |||  
 789 AlaHisValGlnMetAspArgAspAlaValAlaLeuAlaLeuGlnProth 805  
 |||  
 |||  
 3157 GCTCATGCTCTGACGAAAGAAAGAAAGCAAGACCAAGAGCTTGGCCAACT 3206  
 |||  
 |||  
 805 rAspThrIleAsp.....ArgHisValHisGluIleGlnGlnL 818  
 |||  
 |||  
 3207 TGTGGTCTGGAGGCGCCACAGATAAAGAGCTGGAGAACTTGGCCAAA 3256  
 |||  
 |||  
 818 euValLysGluVal.....GluAspLeuGluTyrTrla 828  
 |||  
 |||  
 3257 CCGTGAAGCAACTGTAGAGAAACAGCTGCTTAAGAAAGAAAGAGACCGCA 3306  
 |||  
 |||  
 829 LeuAspSerSerGluValGluValLysSer..... 838  
 |||  
 |||  
 3307 .....TCTGGCTCAGAGAGCCCAATCTGAGCTGCTGCAGCAGCA 3347  
 |||  
 |||  
 839 .....LeuGluGluIleGlnLeuGluLeuAsnPhelG 850  
 |||  
 |||  
 3348 GCCAAGAGCCCAAGCTGGAAGCACTCGGGGCAAGGTAGACAGACTGG 3397  
 |||  
 |||  
 850 LArg..... 851  
 |||  
 |||  
 3398 AACAGCAATGCCAGAACAGCAGAGCAGAGCTGACAGCTTGAAAGCAGC 3447  
 |||  
 |||  
 852 .....ThrArgAspThrLeuIleValG1 859  
 |||  
 |||  
 3448 CTCGAGGCTGAGCGGCTCCCGGGCTGAGCGGAGACAGTCTCTGGAAC 3497  
 |||  
 |||  
 859 uValAsp.....AspLeuArgAspGlnHisArgMetLeuAsnGlnAspMet 875  
 |||  
 |||  
 3498 TCTGCAGGCGCACTTGAAGAGAGAGCCCAAGAGCGTAGAGGCACTGACA 3547  
 |||  
 |||  
 875 er.....SerAlaGln.....Val 879  
 |||  
 |||  
 3548 GTGCTTACCTTCCTGGCCGCAACGGAGACTGGCTGCTCCGACCAAGAGA 3597  
 |||  
 |||  
 880 ArgThrHisAsnAlaArgGluGluLysValLysAlaSerSerIleLeuG1 896  
 |||  
 |||  
 3598 CAAGACCCACGCAAGGCTGAAGCTGAGTGAAGTGAAGGCCCAAGGTGGCCGGG 3647  
 |||  
 |||  
 896 uArgPheGlnLysSerGluGluGluValLeuLeuAlaGluGluLysG 913  
 |||  
 |||  
 3648 CCGGCAAGAGCGTGAAGAGAAAAATAGCTCATACAGCACTTGAGAGAGG 3697  
 |||  
 |||  
 913 InGluIleuIleValGluLysLysLeuLeuGluGluIleuSerLeuAspProLeu 929

```

3698 AGGTGTCATCTGATCCAGGTCTCTGAGAGAGAGGGGAG..... 3741
|||||
930 SerLysGluLysGluSerLeuLeuGlnGlnLysAlaLeuLysGlnLys 946
|||||
3742 AGCAAGAGGTGAAGCGGCTGATGCGCCAGTCAAGAGAGAGCAAGAA 3791
946 sLeuAspGlu..... 949
|||||
3792 GCTGGAGAGAGCTGCGCTCTGACAGCAGAGACAGCCAGCAAGTGC 3841
949 ..... 949
3842 CAGAGCTGCAGACAGCACTGCTCTGCGGAGAGAGTGCAGAGCTCC 3891
950 .....GluTyrHisGlnLeuAlaGluArg.....LysArgG1 960
|||||
3892 GGAAGGAGAGCTGAGAAACAGCGGCTGCTTCAAGAACTGCGCAGAGA 3941
960 uPheGlnGlnLeuAspAlaLeuGlyArgLeuAsnMetLysLysG 977
|||||
3942 GCTGACCTCAGAGCTGAGCTGCGAGAGAGCTGCGCAAGATTGAAG 3991
977 LysTyrLeuAspSer.....LysLysAsnGluLysLeuLysGln 991
|||||
3992 CGTGCAGAGAGAGTCTTCCAGAAAGAGAGCCCTCTCCACCTCGAG 4041
992 GluArgHisVal..... 995
|||||
4042 CTCGAGCAGACACAGACAGCCCTGCTGAGTGAAGTCTGCCAGCTAA 4091
996 ....LeuCysHisSerGlnLeuGlnSerCysMetAlaLysGlnGlnArg 1011
|||||
4092 GCACCTCTGCCAGCAGCTGACGCGCAGAGCCGCTGCCAGAAAGCC 4141
1011 LeuSerAlaGlnLeuAsnLysSerLys..... 1019
|||||
4142 ACCGTGAGAGAGCTGAGAGAGAGAGAGAGCCGCTGCGGAGAGCTCGG 4191
1020 GluLeuLeuGlnGlnGlnGlnLysArgLysAlaLeuAspAsnLe 1036
|||||
4192 GAGCTGCTGCGGCGCCAGCGGAGAGTGGAGAGTATCTCTGCGGCA 4241
1036 uLysTyrArgLysThrLysAlaAspValGlnLeuThrArgAspIle 1053
|||||
4242 GAAAGCTGCAGAGCAGAGCAGAACAGCTCAGAGCTGCGGAGAGAG 4291
1053 LysLeuGlnGlnLysArgLeuSerLysLysLeuSerAlaIleGlu 1069
|||||
4292 CCAGCTATGACAGAGCAGCTGAGCATGCTG..... 4320
1070 AlaAspLeuLysArgHisSerGlnGlnLysGluArgLeuAsnSerGlu 1086
|||||
4321 .....AAGAAGGCCATGCG.....CTGCGGAGAGAGA 4349
1086 eAsnArgTyrGlnGlnLysLeuSerValTyrGlnSerAsnLysSerLys 1103
|||||
4350 GAAACCGGGGCTGCTGAG.....CGGGCCAACTTGGCCCGC 4387
1103 LysLysGlnGlnLeuLysLeuSerGlnTyrLysAspIleGlnLysArg 1119
|||||
4388 ACTTCTCGAATGAGTGGACAGCCCGGAA.....AAGTAT 4428
1120 ThrAsnGlnPheLeuGlnLeuLysThr.....ThrGluMetAlaAs 1133
|||||
4429 GTCCAAAGAGTTGGACCGGTAGCTGATGTCAGACCGCTGCTGCA 4478
1133 nLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMet... 1148
|||||
4479 GGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4528
1149 ArgPheHisSerMetLys.....MetGlnGlnLysnLysIle 1161
|||||

```

```

4529 CCAAGTATGAGAGGTGCCAAGGTCAAGGTCTCTGAGAGAGAGCAGCGGTC 4578
1162 IleLysGlnLeuTyrGlnGlnThrTyrArgGlyGlnAspIleAspTyrI1 1178
|||||
4579 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4628
1178 eSerIleAsnSerAspSerGlnGlyAlaGlyThrArgSerTyrTyr 1195
|||||
4629 ACTGGCTGACTCTGACCAAGCAGCAGAGGTGACAGCAGAGAGAGAGAG 4678
1195 rGValValMetGlnThrGlyAspAlaGluLeuGlnMetArgGlyArgCys 1211
|||||
4679 CTGTCCAGGCTCAGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4728
1212 SerAlaGlnLysValLeuAlaSerLeuIleIleArgLeuAlaLeuAl 1228
|||||
4729 GCCAGCTGAAATGAACTGCAAGCCAGTGTGAGCCAGAGAGAGAGAG 4778
1228 aGluThrPheCysLeuAsnLysGlyIleLeuAlaLeuAspGluProThr 1245
|||||
4779 TGACCACTATTAAGCTGCAG.....ATGAGAGAAAGCCAAA 4813
1245 hAsnLeuAspGlyProAsnAlaGluSer.....LeuAlaAlaLeu 1259
|||||
4814 CACATTTATGATGCCAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4863
1260 LeuArgIleMetGluAlaArgLysGlnGlnLysnPheGlnLeuIleVa 1276
|||||
4864 CGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4902
1276 IleThrHisAspGluArgPheAlaHisLeuIleGlyGlnArgGlnLeu 1293
|||||
4903 CGAGCTGAACTGTAAGCGCTGCGCATGAGCTACAGAGAGAGAGAGAG 4951
1293 LagLys 1295
|||||
4952 AGACCAAG 4959
seq.name: /cgn2_6/prodata/2/lna/5A_COMB.seq: US-08-470-950-3
seq.document: att_block:
; Sequence 3, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOKRATILY, GARY
; TITLE OF INVENTION: LIGAND, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HORNWITZ & THIEBAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,950
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, ESO, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:

```

## SEQUENCE CHARACTERISTICS:

LENGTH: 6306 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

## FEATURE:

NAME/KEY: CDS

LOCATION: 1..6306

PUBLICATION INFORMATION:

AUTHORS: COMPTON, DUANE A

AUTHORS: SZILAK, ILVA

AUTHORS: CLEVELAND, DON W

TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR

TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR

JOURNAL: J. Cell Biol.

VOLUME: 116

PAGES: 1395-1408

DATE: 1992

US-08-470-950-3

## alignment\_scores:

Quality: 376.50

Length: 1486

Ratio: 0.510

Gaps: 62

Percent Similarity: 49.664

Percent Identity: 19.314

## alignment\_block:

US-09-538-396-2 x US-08-470-950-3

Align seg 1/1 to: US-08-470-950-3 from: 1 to: 6306

```
103 glnlysalaserlysmetgluphelaalalegluservalleu... 118
118 .....
802 CAGGGGGCCAGCCACTGGAGCCCAAGAGCTTGAGAGCTGGTACAA 851
851 .....
119 .....
852 GATGAGACCTTACCATGCGCTGATGAMACCTGAGACAGTGCAGG 901
901 .....
130 .. leuserlyargcysalaasmetasparglulerialaleumet 145
145 .....
902 ACCTGAGACAGAGAGAGCCAGATGATCGCAAAATC..... 939
939 .....
146 GlyValSerLysAlaValLeuGlnAsnValIlePheValHisGlnasp 162
162 .....
940 .....
162 userAsnTPProLeuGlnAspProSerThrLeuLysLysPheAsp 179
179 .....
966 CTTTCCTTAAGCTCGGAGATTGCCAGTCACTGCAGCAGCTACAGG 1015
1015 .....
179 spIleheser.....AlaThrArgTyrThrLysAla..... 189
189 .....
1016 ATGCCCTCATGAGCTGACGAGAGAGCCAGCAAGCCACTCAGAGACTGC 1065
1065 .....
190 leuGluValIleLysLysLeuHisLysAsp.....GlnmetGln 203
203 .....
1066 CTAGAGAGAGAGCCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1115
1115 .....
203 uileuserlythPheArgLeuLysLeuGlnAsnLeuGlnThrValLysasp 220
220 .....
1116 CAGAAATATGCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1156
1156 .....
220 InAlaHisLysLeuArgGlnAsnIleAlaGlnAspGlnGlnLysSerAsp 236
236 .....
1157 AACTTTCACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
1206 .....
237 AlaserLysSer.....GlnmetGlnLysLeu 246
246 .....
1207 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1256
```

```
246 sglulysIleCysGlyThrGluArgGluIleLeuGlnmetGluThrSer 263
263 .....
1257 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1306
1306 .....
263 euAspGluLeuArgArgLeuGlnGlnGlnIleAspIleLys...AlaThr 278
278 .....
1307 TAGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
1356 .....
279 GluArg.....SerThrLeuLeu 285
285 .....
1357 GAGGGGGCCACTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1406
1406 .....
285 rglGlnHisGlyLysLeuAlaAlaLeuSerGluGlnAsnGluAspThr 302
302 .....
1407 TAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1456
1456 .....
302 spGluGluLeuMetGluThrGlnThrLysPheGluGluArgGlnLeu 318
318 .....
1457 AGCAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
1506 .....
319 leuGluThrLysIleSerLysLeuValArgAspMetAspArgLys 335
335 .....
1507 CTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1556
1556 .....
335 rtyrSerSerValLeuSerLysGlnAsnSerGluLeuThrHisGluLe 352
352 .....
1557 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606
1606 .....
352 Lys...LeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHis... 366
366 .....
1607 CACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
1656 .....
367 .....
1657 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1706
1706 .....
374 nileCysThrLysHisAsnLeuGlyProValProGlnHisProPheThr 391
391 .....
1707 GGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1735
1735 .....
391 snAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSer 407
407 .....
1736 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1785
1785 .....
408 leuGlnAsn.....AspLeuLeuAspLys 416
416 .....
1786 TTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1835
1835 .....
416 slySerAsnGluAspGlnLeuAspValLeuThrPlysHisTyrLeuLys 433
433 .....
1836 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1879
1879 .....
433 leAsn.....AlaArgTyrSerGluValAspGln... 443
443 .....
1880 CTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1929
1929 .....
444 .....
1930 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1979
1979 .....
456 g.....ArgArgLysAspLysGluLysGluAspAlaAlaGluVal 471
471 .....
1980 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029
2029 .....
471 leuLeuSer...LysPheAsnLeuSerArgIleAspArgLysGluArgHis 486
486 .....
2030 AGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2079
2079 .....
487 MetGlnIleGluValGluArgGlyThrLeuAlaLeuLysGluArgAsp 503
503 .....
2080 GCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
2094 .....
503 rAspSerIleLeuSerGlnLysArgThrGluValTyrSerLeuGlnL 520
```

```
2095 ..... ||| ::::: ::::: |||:::
      ..CAGCTCAGAGCAGCTCCAGAGCCCTCAAGAGT 2128
520 ysileysValleuleuArglulysAspIleleasnarAsnAlaAsp 536
      ::::: |||::: :::::
2129 CTTGAAGGTACACCAAGGCGACGCTTGAAGAGAGAAAGCCAGGCT... 2175
537 GluArgVallylsleuGlyleuLysLysAspAlaLeuGluSerSerLysAs 553
      |||::: |||::: |||:::
2176 ..... GCAGATGCCCTGGAGAGAGCAGCAGCG 2201
553 PLYsleuAsnGluIleValAsnGluHisLysAspLysIleLysVal 570
      ::::: |||::: |||
2202 TTGTATCTGTAGCTGAAGCAGAGACCCGAGAGCTGTGTGAGCAGATA 2251
570 euArgGlyArgAsnProPheGluLysAsp..... 579
      |||::: |||::: |||:::
2252 AGCGGAAACCAAGAGAGCTGMAAGAGAGAGGCTGGCGCCAGGGGCTG 2301
579 ..... 579
2302 GAGGCTGATTACTGACAGCTTGGGAGGCCATCAGAGCTGAGACTGAGT 2351
580 .MetLysLysGluIleAsnGluAlaPhe..... TrpProValAsp 593
      ::::: |||::: |||:::
2352 CCTCGCGCGGAGCTGGCAGAGGCTGCTGCCACACACAGCTGAGA 2401
593 ysgIuTyraAsnGluLeu..... 598
      ::::: |||::: |||:::
2402 GTGAGTGTGACAGCTGCTCAAGAGAGTACGCTGCTGGCGGAGGAT 2451
599 ..ArgSerLysSerGluAlaGluGluGluLysPheThrGlnSe 614
      |||::: |||::: |||:::
2452 GAGATAGCCAGCAGAGAGGAGGACAGTATGGCCCATGTC...CAGA 2498
614 tLysValtLhAspAlaArgGluGluGlnLeuThrLysLeuArgArgAspMet. 630
      ::::: |||::: |||:::
2499 ACAGCTGTGACTTGAAGAGAGATGTGAGAAAGCCCGCAGAGAGCTGC 2548
631 ..AspAlaLysArgArgPheLeuAspSerLysLeuGlnSerIleLeuGln 646
      ::::: |||::: |||:::
2549 AGGAGGCCAAAGAGAGAGGTGGCAGGACATAGATCCACAGGAGACTCCAG 2598
647 Ileser.....AlaAsnValAspMe 653
      |||::: |||::: |||:::
2599 ATAAAGCCGCGCAGACAACTAGTGTGCTCCATGCCAACCTG.... 2643
653 tPheProLysValleuGlnAspAlaMetAsnLysArgAspGluGlnLysA 670
      ::::: |||::: |||:::
2644 ....GCCAGAGCACTCCAGCAGGTCCAGAGAGAGAGAGTCCAGGCCCA 2689
670 tGluGluAsnPheAlaAsnGluMetArgGluMetLeuAlaPro..... 684
      ::::: |||::: |||:::
2690 AGCTTGACGATGACCTCTCCACTGCGAGAAAGATGCTGCCACCAAGC 2739
685 .....PheGluHisLeuAlaArgLys..... 691
      ::::: |||::: |||:::
2740 AAAGAGGTGGCCGCTGTGGAGACCTGTGTCGCAAGCAGAGGTGAGCACA 2789
692 .....AsnHisValCysProCysCysGluArgAlaPheThrP 704
      |||::: |||::: |||:::
2790 GGAAACAGCTCCCGCGGAGTATGTCMAG.....GAGCTCGAGAGGCGAG 2833
704 roAspGluGluAspGluPheValLysLysGlnArg...MetGlnAsnSer 719
      |||::: |||::: |||:::
2834 GAGACAGACAGCCCGATGCTGGAAGCAACAGAGGAGCCAGTTCTGC 2883
720 SerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSerAsnAlaG 736
      |||::: |||::: |||:::
2884 AGCACACAGAGCAGCGCTGACAGCTATGAGCGGAG.....GCAGA 2924
736 uAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyraPheAlaTy 753
      |||::: |||::: |||:::
2925 GCAGATGGCAATGATGATGGAGCGCTGCGG.....G 2956
753 allysLeuValGluGluThrIleProLeuAlaGluLysAsnLeuAsnGln 769
      ::::: |||::: |||:::
2957 CCGCGCTGATGAGAGAGCCAGGCGCAGCAGCAGAGGAGCGCTGGCAGCAG 3006
770 .....HisLeuAlaAspGluSerGlnLysAlaGlnAla... 780
      |||::: |||::: |||:::
3007 GAAAGGAGGTGGCGCGCTGACCCAGAGAGCGGCGCTGCCAGGCTGA 3056
780 ..... 780
3057 CCTTGCCCTGGAGAGAGCGCCAGAGCAGAGCTTGAATGCGCTGCAGA 3106
781 .....PheAspAspLeuGluGlyValLeu 788
      ::::: |||::: |||:::
3107 ACGCCCTCAACAGACAGCGGTGTGTGAGTTGCTTACCTGACCAAGAGCAGCTG 3156
789 .AlaHisValGlnMetAspArgAspAlaValGluAlaLeuGlnProTh 805
      |||::: |||::: |||:::
3157 GCTATGCCCTGACGCGAAGAGAGAGCAGAGCAGAGCTTGGCCAAAGCT 3206
805 AspThrIleAsp.....ArgHisValHisGluIleGlnGln 818
      ::::: |||::: |||:::
3207 TCGTGCTGTGAGGCGACCCAGATAAAGAGCTGAGAGAACTTGGCAAA 3256
818 euValLysGluVal.....GluAspLeuGluTyraAla 828
      |||::: |||::: |||:::
3257 CCGTGAAGCAACTGAGAGAACAGCTGCTTAAAGAAAGAGAGCAGCACA 3306
829 LeuAspSerSerGlyArgGlyValLysSer..... 838
      |||::: |||::: |||:::
3307 .....TCTGCTCAGAGCCCAATCTGAGCTGCTGCTGCAGAGCAGA 3347
839 .....LeuGluGluIleGlnLeuGluLeuAsnPheLeuG 850
      ::::: |||::: |||:::
3348 GCCAAGAGGCCCAAGCTGAGAGCACTGGCGGAGAGGTAGAGCAAGCTGG 3397
850 LnaArg..... 851
      ::::: |||::: |||:::
3398 AACGCAATGCCAGAGAGCAGAGAGCAGCTGACAGCTGAAAGCAGC 3447
852 .....ThraArgAspThrLeuIleValG 859
      |||::: |||::: |||:::
3448 CTCAGAGCTGACGGGCTCCCGGCTGAGCGGACAGTGTCTGGAGAC 3497
859 uValAsp...AspLeuArgAspGlnHisArgMetLeuAsnGluAspMetS 875
      ::::: |||::: |||:::
3498 TCTGCAAGGCGCAGTTAGAGAGAAAGGCCAGAGAGCTAGGCAAGTACA 3547
875 er.....SerAlaGln.....Val 879
      |||::: |||::: |||:::
3548 GTGCTTAGCTCGGCCCAAGCGGAGTGTGCTGCTGCCAGCACAAGTA 3597
880 ArgTrpHisAsnAlaArgGluGluLysValLysAlaSerSerIleLeuG 896
      ::::: |||::: |||:::
3598 CAAGACCAACACAGAGGCTGAAGATGATGGAAGGCCAGGTGGCCGGGG 3647
896 uArgPheGlnLysSerGluGluGluValLeuValleuLeuAlaGluLysG 913
      |||::: |||::: |||:::
3648 CCGGCAAGAGGCTGAGAGAAATAGCTCATCAGCAGCTTGGAGGAG 3697
913 LglGlnLeuIleValGluLysLysLeuLeuGluGluSerLeuAspProLeu 929
      |||::: |||::: |||:::
3698 AGGTGTCCATCTCATGCGCAGAGCTCTGAGAGAGAGAGGAGGAG... 3741
930 SerLysGluLysGluSerLeuLeuGlnGluLysAsnAlaLeuLysGln 946
      |||::: |||::: |||:::
3742 AGCAAGAGATTGAACGGCTGTGTATGGCCAGTACAGAAAGAGCCAGAA 3791
946 sleuAspGlu..... 949
      |||::: |||::: |||:::
3792 GCTGAGAGAGAGCTGCGCTGCTGAGGAGAGAGCAGCAGCAAGTGC 3841
```

949 ..... 949  
3842 CAGAGCTGCAAGCAGCTCTGCTGCGGAGAGAGCTGCGAGCTCC 3891  
950 ..... GluTyrHisGlnLeuAlaGluArg..... LysArgG1 960  
3892 GGGAGGAGGCTGAGAAACAGCGGCTGCTCAGAGAACCTCGCGGAGCA 3941  
960 uPhelGlnGlnLeuAspAlaLeuGlyArgLeuAsnMetLysIleLysG 977  
3942 GCTGACTCAGAGCTGAGCTGCGGAGAGAGCTGCGCCAAAGATTGAAG 3991  
977 LYTyrLeuAspSer..... LysLysAsnGlnLysLeuLysGlnLeuGln 991  
3992 CGTGCGAGGAGAAAGTTCTTCAGAAAGAGCGCCCTCTCCACCTCGAG 4041  
992 G1YArgHisVal..... 995  
4042 CTCGAGCACACACAGAGCCCTGCTGAGTGAAGTCTGCCAGCTAA 4091  
996 ..... LeuCyHisSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgI 1011  
4092 GCAGCTCTGCCAGAGCTGAGCGCGAGAGCGCCGCTGCCAGAAAGCC 4141  
1011 LeSerAlaGlnLeuAsnLysSerLys..... 1019  
4142 ACCGTGAGGAGCTGAGCAGAGCAAGCGCCGCTGGGGGAGCTGGGCA 4191  
1020 GlnLeuLeuGlnGlnGlnGlnGlnLysArgAsnIleAspAspAla 1036  
4192 GAGCTGCTGCGGCGCGAGAGCTGCGGAGCTGATTCCTCTCGCGCA 4241  
1036 uLysTyrArgLysThrLysAlaAspValGlnGlnLeuThrArgAspIleG 1053  
4242 GAGGTGCGCAGAGCGAGCGAGCAAGCTCAGAGCTCGCGGAGAGAG 4291  
1053 LysLeuGlnGlnLysLeuSerGlnTyrLysAspIleGlnLysArgTyr 1069  
4292 CCAGCTATGAGAGCGAGCTGAGCATGCTG..... 4320  
1070 AlaAspLeuLysArgHisSerGlnGlnLysGlnArgLeuAsnSerGln 1086  
4321 ..... AAGAAGCGCCATGCG..... CTGCTGCGCAGAGA 4349  
1086 AsnArgTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1103  
4350 GACCGGCGGCGCGGCTGAG..... CGGGCGCAACCTTGGCGCGC 4387  
1103 LysGlnGlnLeuLysLeuSerGlnTyrLysAspIleGlnLysArgTyr 1119  
4388 AGTTCTGGAAGTGGATTGAGCAGCGCCGCGAA..... AAGTAT 4428  
1120 ThrAsnGlnPheLeuGlnLeuLysThr..... ThrGlnMetAlaAs 1133  
4429 GTCCAGAGAGTTGCGAGCGCTAGCTGATGCTGAGACCCGCTGCTCA 4478  
1133 nLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMet.... 1148  
4479 GGTGACGACGAGAGCAGACAGACAGCTGCCGCGGAGCTGAGATGACG 4528  
4529 CCAAGTATGAGGTGCCAAGGTCAAGCTCTGAGAGAGAGCGCGCTTC 4578  
1149 .. ArgPheHisSerMetLys..... MetGlnGlnLeuAsnLysIle 1161  
1162 IleLysGlnLeuTyrPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1178  
4579 CAGGAGAGAGCGAGAACTCACTGCGCAGGTGAGAACTGAGTAGAA 4628  
1178 eSerIleAsnSerAspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1195  
4629 ACTGCTGACTCTGACCAAGCAGAGAGGTGCGAGCAGAGAACTGAGAG 4678

1195 rGValValMetGlnThrGlyAspAlaGlnLeuGlnMetArgGlyArgCys 1211  
4679 CTGTCCAGGCTCAGGAGGAGGAGAGCGAGCGCCAGCGCTTCAG 4728  
1212 SerAlaGlnLysValLeuAlaSerLeuIleLeuArgLeuAlaLeuAl 1228  
4729 GCCAGCTGAGTGAATCCAGCCAGTTCAGCCAGAGAGAGCGAGC 4778  
1228 aGlnThrPheCysLeuAsnGlyLeuLeuAlaLeuAspGlnProTyrT 1245  
4779 TGAGCACATTAAGCTGCG..... ATGAGAGAAACCAAAA 4813  
1245 hrAsnLeuAspGlyProAsnAlaGlnSer..... LeuAlaAlaLeu 1259  
4814 CACATTATGATGCCAAGAGCAGAGCAAGCAAGAGCTGAGAGAGCTG 4863  
1260 LeuArgIleMetGlnAlaArgLysGlnGlnGlnGlnGlnGlnGlnGln 1276  
4864 CGAGCCTGAGAGCTGAGAG..... GAAACAAAGAGCTG.... 4902  
1276 IleThrHisAspGlnArgPheAlaHisLeuIleGlyGlnArgGlnLeu 1293  
4903 ..CGAGCTAAGCTGAACGCTGCGCATGAGCTACAGCAGAGCTGAGCTGA 4951  
1293 LysLys 1295  
4952 AGACCAAG 4959  
seq\_name: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:US-08-467-781-3  
seq\_documentation\_block:  
; Sequence 3, Application US/08467781  
; Patent No. 5780596  
GENERAL INFORMATION:  
APPLICANT: TDKRATLY, GARY  
APPLICANT: LIDARD, GRAHAM P  
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESS: TESTA, HURWITZ & THIBAUDT  
STREET: 125 HIGH STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,781  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER, ESO, EDMUND R  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: MTP-013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7100  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6306 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..6306  
PUBLICATION INFORMATION:

AUTHORS: COMPTON, DUANE A  
 AUTHORS: SZILAK, ILVA  
 AUTHORS: CLEVELAND, DON W  
 TITLE: PRIMARY STRUCTURE OF NDMA, AN INTRANUCLEAR  
 TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR  
 TITLE: SEGREGATION OF PROTEINS AT MITOSIS  
 JOURNAL: J. Cell Biol.  
 VOLUME: 116  
 PAGES: 1395-1408  
 DATE: 1992  
 US-08-467-781-3

US-08-467-781-3

alignment scores:  
 Quality: 376.50 Length: 1486  
 Ratio: 0.510 Gaps: 62  
 Percent Similarity: 49.664 Percent Identity: 19.314

alignment block:  
 US-09-538-396-2 x US-08-467-781-3 ..

Align seg 1/1 to: US-08-467-781-3 from: 1 to: 6306

103 GlnIyAlaSerLysMetGluPhelYsAlaIleGluSerValLeuGln.. 118  
 802 CAGGGCCGACCCCATGAGCCAGAGAGCTTGAGAGCTGCTGACAA 851  
 119 .....ThrlLeaSProHlsThrlGlyGluLysValCys... 129  
 852 GAATGAGAGCTTACATGGCGGCTGCATGAACCTGAGAGAGCGCAGG 901  
 130 ..LeuSerTYrArgCysAlaAspMetAspArgGluIleProAlaLeuMet 145  
 902 ACCTGAG 939  
 146 GlyAlaSerLysAlaValLeuGluAsnValIlePheValHisGlnAspG 162  
 940 .....AACCACTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965  
 162 UserAsnTYrProLeuGlnAspProSerThrlLeuLysLysPheAsp 179  
 966 CCTTCCTTAAAGTGGGGAGTGGCAGATCATCTGACAGAGCTACAGG 1015  
 179 spllePheSer.....AlaThrArgTYrThrlYsAla..... 189  
 1016 ATGCCCTCAATGAGCTGACGAGAGAGAGAGAGAGAGAGAGAGAG 1065  
 190 LeuGluValIleLysLysLeuHisLysAsp.....GlnMetGlnG 203  
 1066 CTAG 1115  
 203 uIlLeuThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAsp 220  
 1116 CAAGAAATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1156  
 220 InAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnGluLysSerAsp 236  
 1157 AACCTTCACAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206  
 237 AlaSerLysSer.....GlnMetGluGlnLeuLys 246  
 1207 CAG 1256  
 246 sGluLysIleCysGlyThrlArgGluIleLeuGlnMetGluThrSerL 263  
 1257 GCAG 1306  
 263 euAspGluLeuArgArgLeuGlnGlnIleAspIleLys..AlaThr 278  
 1307 TAGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356  
 279 GluArg.....SerThrlLeuLeuTh 285

1357 GAGCGGGGCCACTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1406  
 285 rGlnGlnHisGluLysLeuAlaIleuSerGluGluAsnGluAspThra 302  
 1407 TGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1456  
 302 sPlGluLysMetGluTYrPglThrlLysPheGluGluArgIleAlaLeu 318  
 1457 AGCAGGCTTCCAGGCTCATGGGGCCCGTTGACTGCTCCAGAGTGGCT 1506  
 319 LeuGluThrlLysIleSerLysLeuValArgAspMetAspAspGluAla 335  
 1507 CTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1556  
 335 rTYrSerSerValLeuSerLysGlnAsnSerGluLeuThrlHisGluIle 352  
 1557 AGAAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606  
 352 LysLys..LeuGlnAlaGluAlaAspAlaHisLeuThrlMetLysHis 366  
 1607 CACAG 1656  
 367 .....GluArgAspSerAspIleLysAs 374  
 1657 GTGAG 1706  
 374 nIleCysThrlLysHisAsnLeuGluProValProGluHisProPheThra 391  
 1707 GGTAGCGGAGAGAG.....CAGAGAGAGAGAGAGAGAGAGAGAG 1735  
 391 snAspValAlaMetAsnLeuThraAsnArgIleLysAlaArgLeuSerSer 407  
 1736 AGGAG 1785  
 408 LeuGluAsn.....AspLeuLeuAspLysLys 416  
 1786 TTAAG 1835  
 416 LysSerAsnGluAspGlnLeuAspValLeuThrlPylHisTYrLeuLysI 433  
 1836 GAAGGCT.....GCCAAGCTGAGATTCGACAGAGAGAGAGAGAG 1879  
 433 LeuAsn.....AlaArgTYrSerGluValAspGluLys 443  
 1880 CTATAGTACGCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1929  
 444 .....IleGlnSerLysIleGluSerMetSerGlyIleLeuArg 456  
 1930 GAG 1979  
 456 g.....ArgArgLysAspLysGluLysGluArgAspAlaGluValG 471  
 1980 GACAG 2029  
 471 LuLeuSer..LysPheAsnLeuSerArgIleAspGluArgGluArgHis 486  
 2030 AGTTCAG 2079  
 487 MetGlnIleGluValGluArgLysThrlLeuAlaLeuGluLysArgAsp 503  
 2080 GCCAGG.....GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094  
 503 rAspSerIleLeuSerGluLysArgThrlGluValTYrSerLeuGluL 520  
 2095 .....CAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2128  
 520 yslIleLysValLeuLeuArgGluLysAspIleLeuAsnArgAspAlaAsp 536  
 2129 CTTGAAGGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2175  
 537 GluArgValLysLeuGluLysLysAspAlaLeuGluSerLysAs 553  
 2176 .....GCAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2201

553 pLYLeuasnGLuIleValaEnGLuHISLysAspLYSleuLYSValL 570  
2202 TTGATCTCTAGCTGAGAGCAGAGACCCGAAAGCTGTGTGAGCAGCAT 2251  
570 euArgGLyArgAspProPheGLuLYSAsp 579  
2252 AGCGGGAACGAAAGAGCTGAGAGAGAGAGGCTGGCGCAAGGGCTG 2301  
579 ..... 579  
2302 GAGCTCGATTACTGACAGCTTGGGAGGCCCATCAGCTGAGACTGAAGT 2351  
580 .MetLYSArgGLuIleasnGLuAlaPhe.....TrpProValAspL 593  
2352 CCGCGGCGGAGAGCTGGCAGAGGCCATGCTGCCAGCAGCAGCTGAGA 2401  
593 YSGluTYrAsnGLuLeu..... 598  
2402 GTGAGCTGAGCAGCTCTCAAGAAGTAGCTGCTGGCGCTGACGGGTAT 2451  
599 ..ArgSerLYSArgGLuAlaGLuIleuLYSPhetHGLnSe 614  
2452 GAGGATAGCCAGCAGAGAGAGGACAGATAGCGCCATGTTTC...CAGA 2498  
614 rLYSValThrAspAlaArgGLuIleuThrLYSLeuArgArgAspMet. 630  
2499 ACAGCTGATGACTTGAAGAGAGAAATGTAGAGAGGCCCGCCAGAGCTCG 2548  
631 ..AspAlaLYSArgArgPheLeuAspSerLYSLeuInSerIleuGLn 646  
2549 AGGAGCGAAAGAGAGAGCTGGCAGGCAATGCCACAGCGAGCTCCAG 2598  
647 IleSer.....AlaAsnValAspMe 653  
2599 ATAAAGCCGCGCAGACAACTAGCTGAGCTCCATGCCCAACCTG..... 2643  
653 tHeProLYSValLeuGLnAspAlaMetAsnLYSArgAspGLuIleLYS 670  
2644 ...GCCAAGACACTCCAGCAGAGTCCAGAGAGAGAAAGTCAAGGCCAG 2689  
670 rGLnLeuAsnPheAlaAsnGLyMetArgLumetLeuAlaPro..... 684  
2690 AGCTTGACATGACTCTCCACTCTCCAGAAAGATGGCTGCCACACAGC 2739  
685 .....PheGLuHISLeuAlaArgLYS..... 691  
2740 AAAGAGCTGGCCCGCTTGGAGACCTTGTCCCGCAGCAGAGGTGAGCAGCA 2789  
692 .....AsnHISValCysProCysGLuArgAlaPheThr 704  
2790 GGAAACAGCTCCCGGAGTGTAGTCAAG.....GAGCTGCGAGGGCAG 2833  
704 roAspGLuIleAspGLuPheValLYSGLnArg...MetGLnAsnSer 719  
2834 GAGACAGACAGCCCGAGTGGCTGGAAGAGCAGACAGGAGCCAGTTCTGC 2883  
720 SerThrAlaGLuArgSerLYSAlaLeuAlaMetGLuSerSerAsnAlaGL 736  
2884 AGCAGACAGCAGCGCTGAGCTATGAGCGGGAG.....GCAGA 2924  
736 uAlaLeuPheGLnGLuLeuAspLYSLeuArgThrIleTYrAspAlaArgY 753  
2925 GCAGATGGCGCAATGAGCTGGAACGGCTGCG.....G 2956  
753 alLYSLeuValGLuIleuThrIleProLeuAlaGLuLYSAsnLeuAsnGLn 769  
2957 CCGCGCTGATGAGAGCCAGGCGCAGCAGCAGAGAGAGCGCTGGCAGCAG 3006  
770 .....HISLeuAlaAspGLuSerGLnLYSAlaGLnAla.. 780  
3007 GAAAGGAGAGCTGGCGGCTGAGCCAGAGCGGGCGCTGCCAGAGCTGA 3056  
780 ..... 780  
3057 CCTGCCCTGGAGAGAGCGGCCAGAGCAGAGCTTAGTCCGGCTGCAGA 3106  
781 .....PheAspAspLeuLeuGLyValLeu 788  
3107 AGCCCTCAGAGAGAGCGGTGGAGGTTCCGACACCTCCAGAGAGCAGCTG 3156  
789 AlaHISValGLnMetAspArgAspAlaValAlaLeuLeuGLnProth 805  
3157 GCTCATGGCCCTAGCGAAAGAGAGCAGAGCAGAGATTGGCCCAAGCT 3206  
805 rAspThrIleAsp.....ArgHISValHISGLuIleGLnGLnL 818  
3207 TCGTGATCTGGAGCGCAGCCAGATAAAGAGCTGAGAACTCGGCAAA 3256  
818 euValLYSLeuVal.....GLuAspLeuGLuTYrAla 828  
3257 CCGTAGAGCACTGAAAGAGACAGCTGGCTTAAGAAAGAAAGAGAGCAGCA 3306  
829 LeuAspSerSerGLyArgGLyValLYSArg..... 838  
3307 .....TrpGCGTCAAGAGCCCAATCTGAGCTGTGGCAGAGACAGA 3347  
839 .....LeuGLuIleGLnLeuGLuIleuAsnPheLeuG 850  
3348 GCCAAGAGCCCGCAAGCTGGAAGACTGGCGGAGAGGTGAGCAAGCTGG 3397  
850 LArg..... 851  
3398 AACAGCAATGCCAAGAACAGCAGAGCAGAGCTGACAGCTCGAAGCCAGC 3447  
952 .....ThrArgAspThrLeuIleValGL 859  
3448 CTGAGAGCTGAGAGCGGCTCCCGGGGTGAGCGGAGACAGTCTGTGAGAC 3497  
859 uValAsp...AspLeuArgAspGLnHISArgmetLeuAsnGLuAspMetS 875  
3498 TCTGACAGGCGCACTGAGAGAGAAAGCCAGAGAGTGAAGCAGACAGTCA 3547  
875 er.....SerAlaGLn.....Val 879  
880 ArgThrHISAsnAlaArgGLuIleuLYSValLYSAlaSerSerIleuGLn 896  
3598 CAAGACCAAGCAGAAAGCTGAAATGAGTGAAGCCAGCTGGCCGGGG 3647  
896 uArgPheGLnLYSArgGLuIleuLeuValLeuAlaGLuIleuLYS 913  
3648 CCGGCAAGAGCTGAGAGAGAAATAGCCATCAGCAGAGCTTGAAGAGAG 3697  
913 LuGLnLeuIleValGLuLYSLeuLeuGLuIleuLeuAspProLeu 929  
3698 AGGTGTCATCTGAAATCCGAGAGCTCTGGAAGAGAGGGGGAG..... 3741  
930 SerLYSGLuLYSArgLeuLeuGLnGLuTYrAsnAlaLeuLYSGLnLY 946  
3742 AGCAAGAGAGTTGAAGCGCTGTGATGCGGAGTCAAGAGAAAGCCAGAA 3791  
946 sLeuAspGLu..... 949  
3792 GCTGAGAGAGAGCTGCCCTGTCTGACAGCAGAGACAGCCAGCAAGCTGC 3841  
949 ..... 949  
3842 CAGAGCTGCAAGAACGAGCTCTGCTGCGGAGAGAGCTGACAGCTCC 3891  
950 .....GluTYrHISGLnLeuAlaGLuArg.....LYSArgGL 960  
3892 GGGAGGAGAGCTGGAAGAAAGAGCGGTGCTTCAAGAAACCTGGCGAGGA 3941  
960 uPheGLnGLuIleuAspAlaLeuGLyArgLeuAsnMetLYSleuLYS 977



```

3942 GCTGACCTCAGACGCTGAGCTGGGAGAGACCTGGCCAAAGATTGAAG 3991
977 LTYTLEuAspSer....LysLysasnGLuLysLeuLysGLuLeuGLn 991
3992 CGTGACGAGAAAGTTCTTCCAGAAAGAGCGCTCTCCACCTCGAG 4041
992 GLYArgHisVal..... 995
4042 CTCGAGCACACACAGACAGCCCTGGTGAGTAGAGCTGCTCCAGCTAA 4091
996 ....LeuCyHisSerGLuLeuGLnSerCyMetAlaLysGLnArgI 1011
4092 GCACCTCTGCACAGAGCTGACGAGCCGAGAGCGCTGCTCCGAAAGCC 4141
1011 LeSerAlaGLuLeuAsnLysSerLys..... 1019
4142 ACCGTGAGGAGCTGGAGCAGAGCAGAGCCGCTGGGGAGCTCGGGCA 4191
1020 GLuLeuLeuGLnGLnGLnGLnGLnLeuLysArgAsnLeuAspAsnLe 1036
4192 GAGCTGCTGCGGGCCGAGGAGCTGGGAGCTGATCTCTGCGGCA 4241
1036 ULysTyrArgLysThrLysAlaAspValGLuGLnLeuThrArgAspLe 1053
4242 GAGGTGCGACAGAGCAGAGCAGAGCTCAGACAGCTGCGGGCAGAGAG 4291
1053 LysSerLeuGLuArgLeuLeuSerLysLysSerLeuSerAlaLeu 1069
4292 CCAGCTATGACAGAGCAGCTGAGCTG..... 4330
1070 AlaAspLeuLysArgHisSerGLnGLnLysGLuArgLeuAsnSerGLuPh 1086
4331 .....AAGAGCGCGATGCG.....CTGCTGCGAGAGCA 4349
1086 eAsnArgTyrPrlngLThrLeuSerValTyrGLnSerAsnLysSerLys 1103
4350 GAGCGGGGGCTGGGTAG.....CGGGCCAACTTGGCGCGC 4387
1103 LsLysGLnGLuLeuLysLeuSerGLnTyrLysAspLysArgTyr 1119
4388 AGTTCTGGAAGTGGAGTTGACAGCGCGCGAA.....AAGTAT 4428
1120 ThrArgnLpLeuGLnLeuLysThr.....ThrgLMetAlaAs 1133
4429 GTCCAGAGTTGGCAGCCGTACGTGATGCTGAGACCCCTGCTGCTGA 4478
1133 nLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMet... 1148
4479 GGTGACGAGGAGAGCAGAGCAGCTGCGGGAGCTGAGTATGACTG 4528
1149 ..ArgPHisSerMetLys.....MetGLuLysLysLysLys 1161
4529 CCAAGTATGAGGTCGCAAGGTCAAGTCTGAGAGAGCAGCGGCTC 4578
1162 LLeuLysGLuLeuTyrPrlngLThrTyrArgLysGLnAspLysLys 1178
4579 CAGGAGAGAGGAGAGAACTGCTGCCAGGTGGAGAGACTGAGAGAGAA 4628
1178 eSerLysAsnSerLysSerGLuLysAlaGLnTyrArgSerTyrSerTyr 1195
4629 ACTGCTGACTGCTGACCAAGCAGAGAGTGCAGAGCAGAGAGAGAG 4678
1195 rGValValMetGLnThrLysAspAlaGLuLeuGLuMetArgLys 1211
4679 CTGTGACGAGCTGCGAGGAGCAGAGCCAGAGAGCCAGCTTCAG 4728
1212 SerAlaGLnLysValLeuAlaSerLeuLeuLeuArgLeuAlaLeu 1228
4729 GCCCAGCTGATGAAGCTGCAAGCCAGTTGAGCCAGAGAGAGAGAGC 4778
1228 aGLuThrPheCyLeuAsnCyGLyLysLeuAlaLeuAspLysLys 1245

```

```

4779 TGACCACTATAAGCTCGAG.....ATGGAGAAAGCCAAA 4813
1245 hAsnLeuAspGLyProAsnAlaLeuSer.....LeuAlaAlaLeu 1259
4814 CACATATGATGCCAAGAGAGCAGAGCAAGCAAGAGCTGCAGAGAGAGCTG 4863
1260 LeuArgLysMetGLuAlaArgLysGLyGLnLysAsnPheGLnLeuLeu 1276
4864 CGGAGCTGAGAGCTGACAGAG.....GAAACAAAGAGCTG..... 4902
1276 LLeuThrHisAspLysArgPheAlaHisLeuLeuLeuGLnArgLysLeu 1293
4903 CGAGCTGAAAGCTGAGCGGCTGCGGCAATGAGCTACAGAGAGCTGCTGA 4951
1293 LaGLuLys 1295
4952 AGACCAAG 4959

seq_name: /c9n2_6/prodata/2/lna/5B_COMB.seq:US-08-483-924-3
seq_documentation block:
; Sequence 3, Application US/08483924
; Patent No. 5882876
; GENERAL INFORMATION:
; APPLICANT: TOOKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,924
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDWARD R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:
; AUTHORS: COMPTON, DUANE A
; AUTHORS: SZILAK, ILYA
; TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
; JOURNAL: J. Cell Biol.
; VOLUME: 116
; PAGES: 1395-1408
; DATE: 1992

```

US-08-483-924-3

```
alignment_scores:
  quality: 376.50      Length: 1486
  ratio: 0.510         Gaps: 62
Percent Similarity: 49.664      Percent Identity: 19.314
```

```
alignment_block: 1
ms-00-539-306-3 115-08-483-924-3
```

Align seg 1/1 to: US-08-483-924-3 from: 1 to: 6306

[illegible]

319 LeuGluThrLysIleSerLysLeuValArgAspMetHisArgGluIle 335  
 1507 CTGACCTTGAGCTCCACACACATAATGCCACCATTCACGACACAGATCA 1556  
 335 rTysSerSerValLeuSerLysGlnAspSerGluLeuThrHisGluIleG 352  
 1557 AGAACTGGCTGGCTTGAAGCAGCGCCAAAGAAAGAAACACAGGCCAGTAA 1606  
 352 LysLys... LeuGlnAlaGluIleAspAlaHisLeuThrMetLysHis... 366  
 1607 CACAAACCTCCAAACAGAAAGACAGGGCTCCAGAGGCTCTCCGCCACCAAG 1656  
 367 ..... GlnArgAspSerAspIleLysAs 374  
 1657 GTGAGACAGCTAAAGCAGTAAAGCTGAAAGAGAAAGAGACGCGTTGAAGA 1706  
 374 nLecArgThrLysHisAsnLeuGlyProValProGluHisProPheThr 391  
 1707 GGTACCGGGAAG..... CAGAGGCAACTAGGC 1735  
 391 snAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSerSer 407  
 1736 AGACACCAATCCACAGAACTGGCACTGTCGAGAGAGACGAGAGGCTCTCC 1785  
 408 LeuGluAsn..... AspLeuLeuAspLysLys 416  
 1786 TTAAGGAGCCGGAGAGCGGCTCTCAAGCAGCTGAGCGCCATGGAGAAGA 1835  
 416 sLysSerAsnGluAspGlnLeuAspValLeuThrLysHisArgLysLysI 433  
 1836 GAAGGCT..... GCCAAGCTGAGATTCCTGACGACCACTTACAGTGG 1879  
 433 LeAsn..... AlaArgTyrSerGluValAspGlyGln... 443  
 1880 CTAATCAACCCCGGAGACAGTGGCCAGACTGATGATACACAGGCCACAGGG 1929  
 444 ..... IleGlnSerLysIleGluSerMetSerGlyIleLeuArg 456  
 1930 GAGAAGCAGACCTGACCGCGAAGGTGGAGGAAATCCAGGCTGTGTGA 1979  
 456 g..... ArgArgLysAspLysGlnLysGlnArgAspAlaIleGluValG 471  
 1980 GACAGCCGCCAGGAACACATGAGGCCACAGGCCACAGTTCGACAGCTAG 2029  
 471 IuLeuSer... LysPheAsnLeuSerArgIleAspGluArgGluArgHis 486  
 2030 AGTTGCAGCTGGCGTGTGACGACAAAGAACCAATCGAGAAAGAGGGTGG 2079  
 487 MetGlnIleGluValGluArgLysThrLeuAlaLeuGluGlyIuArgAspTy 503  
 2080 GCCCAG..... GAAAGAGAC.. 2094  
 503 rasPSerIleIleSerGlnLysArgThcGluValTyrSerLeuGluGlnL 520  
 2095 ..... CACTCCAGAGCAGCTCCAGGCTCTCAAGAGT 2128  
 520 ystIleLysValLeuLeuArgGluLysAspIleIleAsnArgAsnAlaAsp 536  
 2129 CCTTGAAGGTCAACCAAGGCGACGCTTGAAGAGAGAAAGCGCAGGCT... 2175  
 537 GluArgValLysLeuGlyLeuLysLysAspAlaLeuGluSerSerLysAs 553  
 2176 ..... GCAGATGCCCTTGAGAGACACACACCG 2201  
 553 pLysLeuAsnGluIleValAsnGluHisLysAspLysIleLysLysValL 570  
 2202 TTGTATCTTCGACCTTAAGGCAAGACCCGAAAGCTGTGTGGAGACATAT 2251  
 570 euArgLysArgAsnProPheGluLysAsp..... 579  
 2252 AGCGGGAACCAAGAGACTGGAAGAAAGACAGGCTGGGGCGCAAGGGCGTG 2301  
 579 ..... 579

2302 GAGGCTCATTACTGACGCTTGGGGAGGCCCATCAGGCTGAGACTGAAAGT 2351  
580 .MetLysLysGluIleAsnGlnAlaPhe.....TrrProValAspL 593  
2352 CCGCGCGGCGGAGCTGGCGAGGCGCATGGCTGCCAGCACACAGCTGAGA 2401  
593 yGluIuYrAsnGluLeu..... 598  
2402 GTGAGCTGAGCAGCTCTCAAGAAAGTACTGCTGGCGGTGACGGGAT 2451  
599 ...ArgSerLysSerGlnGlnIleProLeuAlaGluLysAsnIleuGln 614  
2452 GAGGATAGCGACAGAGAGAGGAGGACAGTATGGCGCATGTTG...CAGA 2498  
614 rLyValThrAspAlaArgGluGlnLeuThrLysLeuArgAspMet. 630  
2499 ACAGCTGATGACTTTGAGAGAGGAATGTGAGAAAGCCCGCCAGAGGCTGC 2548  
631 ..AspAlaLysArgArgPheLeuAspSerLysLeuGlnSerIleuGln 646  
2549 AGGAGCCAAAGAGAGAGGTGGCAGGCAATGATCCACAGCAGGACTCAG 2598  
647 IleSer.....AlaAsnValAspMet 653  
2599 ATAGCCGCGCAGAGAAACAAGTACTGAGTCTCATGCCAACCTG.... 2643  
653 tPheProLysValLeuGlnAspAlaMetAsnLysArgAspGluGlnLysA 670  
2644 ....GCCAAGCACTCCACAGGCTCCAAAGAGAGAAAGTCAAGGCCCA 2689  
670 rGluGluAsnPheAlaAsnGluMetArgGluMetLeuAlaPro..... 684  
2690 AGCTTCATATGACTCTCTCTCAGGAAAGATGGCTGCCACACAGC 2739  
685 .....PheGluIleLeuAlaArgLys..... 691  
2740 AAGAGGTGGCCGCTGGAGACCTGTGCCGAAAGGAGGAGTGCAGCAGCA 2789  
692 .....AsnIleValCysProCysGluArgAlaPheIleP 704  
2790 GGAACAGGCTCCCGGAGATTAGTCAAG.....GAGCTCGAGAGGCCAG 2833  
704 rGlyAspGluAspArgPheValLysLysGlnArg...MetGlnAsnSer 719  
2834 GAGACAGACAGCCCGAGTGGCTGGAGAGCAACAGGACGCCAGTTCTGC 2883  
720 SerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSerAsnAlaG 736  
2884 AGCAGACAGCAGGCTGCAGGCTATGAGCGAGAG.....GCAGA 2924  
736 uAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyAspAlaTy 753  
2925 GCAGATGGGCAGATGAGCTGAAAGGCTGCGG.....G 2956  
753 alLysLeuValGluGluThrIleProLeuAlaGluLysAsnLeuGln 769  
2957 CCGCGGTGATGGAGAGCCAGAGGAGCAGCAGAGAGGAGCGTGGCGAGCG 3006  
770 .....HisLeuAlaAspGluSerGlnLysAlaGlnAla.. 780  
3007 GAAGGAGAGGTGGCGGCTGACCCAGAGCGGGCGGCTGCCAGAGCTGA 3056  
780 ..... 780  
3057 CCTGGCCCTGGAGAGAGCGGCCAGAGCAGAGCTTGATGCGCGCTGCAGA 3106  
781 .....PheAspAspLeuLeuGluValLeu 788  
3107 AGCGCCTCAACGAGCAGCGTGTGGAGTTCGATCCACTGCAGAGAGGACTG 3156  
789 AlaHisValGlnMetAspArgAspAlaValGluAlaLeuGlnProth 805

805 rAspThrIleAsp.....ArgHisValHisGluIleGlnGln 818  
3207 TCGTGGCTGTGGAGGACGCCCGAGATAAAGACTTGAGAGACTTGGGAAA 3256  
818 euValLysGluVal.....GluAspLeuGluTyAla 828  
3257 CCGTGAAGCACTGAAAGAACAGCTGGGTTAGAAAGAAAGAGCAGCAGCA 3306  
829 LeuAspSerSerGlyArgGlyValLysSer..... 838  
3307 .....TCTGGCTACAGGCCCACTGTAGAGCTGCTGGCAGAGCAGA 3347  
839 .....LeuGluGluIleGlnLeuGluLeuAsnPheLeuG 850  
3348 GCCAAGAGGCCCAAGCTTGGAGAGCACTGGCGGCGAGAGGTGAGCAAGCTG 3397  
850 LArg..... 851  
3398 AACAGCAATGCCAGAGAGCAGAGCAGGCTGACAGCTGGAACGAGC 3447  
852 .....ThrArgAspThrIleValG 859  
3448 CTCGAGCTGAGCGGGCTCCGGGCTGAGAGGAGCAAGTCTGTGAGAC 3497  
859 uValAsp...AspLeuArgAspGlnHisArgMetLeuAsnGluAspMet 875  
3498 TCTGAGGGCCAGTTAGAGAGAAAGCCACAGACTAGGCGACAGTCA 3547  
875 er.....SerAlaGln.....Val 879  
3548 GTGCTTGTAGCTCGGCCCAAGGAGTGTGCTGCTTCCGACCAAGGTA 3597  
880 ArgThrPheAsnAlaArgGluGlnLysValLysAlaSerIleLeuG 896  
3598 CAAGACCCACACAGAGCTGAAGATGAGTGAAGGCCCGGAGTGGCCGGG 3647  
896 uArgPheGlnLysSerGluGluGlnLeuValLeuLeuAlaGluGluLysG 913  
3648 CCGGCAAGAGCTGAGAGAAATAGCTCATGACAGCAGCTTGAGAGAGG 3697  
913 LgluLeuIleValGlnLysLysLeuLeuGluLysLeuAspProLeu 929  
3698 AGGTCTCATCTGCTGATGCGCAGGCTCCGAGAGAGAGGAGGAG.... 3741  
930 SerLysGluLysGluSerLeuLeuGlnGluLysAsnAlaLeuLysGln 946  
3742 AGCAAGGAGTTGAAGCGGCTGTATGGCCAGATCAGAAAGACCCAGAA 3791  
946 sLeuAspGlu..... 949  
3792 GCTGAGAGAGAGCTGCGCTCTCTGACAGCAGACAGCCACACAGCTGC 3841  
949 ..... 949  
3842 CAGAGCTCAGAACGCAAGCTGCTGCGGAGAGAGTGCAGAGCTTC 3891  
950 .....GluTyHisGlnLeuAlaLysArg..LysArgG 960  
3892 GAGAGGAGGCTGAAACACAGCGGTGCTTCAGAGAACTGCGGCGAGGA 3941  
960 uPheGlnGlnLeuAspAlaLeuGlyArgLeuAsnMetLysIleLysG 977  
3942 GCTGACTCTACAGGCTGAGCGTGGGAGAGAGCTGGGCCCAAGATTGAAG 3991  
977 LTyTyLeuAspSer.....LysLysAsnGlnLysLeuLysGlnGln 991  
3992 CGTGGCAGAGAAAGTTCTTCAGAAAGAGAGGCCCTTCACACCTCGAG 4041  
992 GlyArgHisVal..... 995  
4042 CTCGAGCACACAGCAGCAGGCTGCTGATGATGAGTGCCTCCAGCTAA 4091

```

996 ...LeucylHisSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgI 1011
1011 leserAlaGlnLeuAsnLysSerLys..... 1019
4142 ACCGTGAGAGAGCTGAGACAGACAGAGCCGCTGGGGAGACTGCGGCA 4191
4092 GCACCTCTGCCAGCTGCAGAGCCCGAGAGCGCTGCCGAGAAAGCC 4141
1011 leserAlaGlnLeuAsnLysSerLys..... 1019
1020 GlnLeuLeuGlnGlnGlnGlnGlnLysArgAsnLysAspAsnLe 1036
4192 GACCTGCTGGGGCCCGAGAGCTGGAGAGCTGATCTCTCGCGCA 4241
1036 ulysTyrArgLysThrLysAlaAspValGlnGlnLeuThrArgAspLeu 1053
4242 GAAGGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4291
1053 luserLeuGlnLysArgLysLeuSerLysLeuSerLysLeuSerLys 1069
4292 CCAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320
1070 AlaAspLeuLysArgHisSerGlnGlnLysGlnLysLysAsnSerGluPh 1086
4321 ...AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4349
1086 eAsnArgTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1103
4350 GAACCGGGCGCTGGTGAG.....CGGGCCAACTGCGCGC 4387
1103 lslsGlnGlnLysLeuSerGlnTyrLysAspLeuGlnLysArgTyr 1119
4388 AGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4428
1120 ThrAsnGlnPheLeuGlnLysThr.....ThrgLysMetAlaAs 1133
4429 GTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4478
1133 nLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMet... 1148
4479 GGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4528
1149 ...ArgPheHisSerMetLys.....MetGlnLysLysAlaAsnLys 1161
4529 CCAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4578
1162 lLysGlnLeuTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1178
4579 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4628
1178 eSerLysAsnSerAspSerGlnGlnGlnGlnGlnGlnGlnGlnGln 1195
4629 ACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4678
1195 rGValValMetGlnThrGlnAspAlaGlnLysGlnLysGlnLysGln 1211
4679 CTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4728
1212 SerAlaGlnGlnLysValLeuAlaSerLeuLeuLeuArgLeuAlaLeu 1228
4729 GCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4778
1228 agLThrPheCysLeuAsnGlyLysLysLysLysLysLysLysLys 1245
4779 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4813
1245 hrAsnLeuAspGlyPyrAsnAlaGlnUser.....LeuAlaAlaAlaLeu 1259
4814 CACATTATGATGCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4863
1260 LeuArgLysMetGlnLysLysGlnGlnGlnGlnGlnGlnGlnGlnGln 1276
4864 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4902

```

```

1276 lLysThrHisAspGlnLysArgPheAlaHisLeuLeuGlnGlnArgGlnLeu 1293
4903 CGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4951
1293 lArgLys 1295
4952 AGACCAAG 4959

```

seq\_name: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:US-08-195-487-3

seq\_documentation\_block:

Sequence 3, Application US/08195487

Patent No. 5783403

GENERAL INFORMATION:

APPLICANT: TOURKATLY, GARY

TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESSES:

ADDRESS: TESTA HURWITZ & THIBEAULT

STREET: 53 STATE STREET

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/195,487

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/901,701

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: PITCHER ESQ, EDMUND R

REGISTRATION NUMBER: 27,829

REFERENCE/DOCKET NUMBER: MTP-013

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/248-7000

TELEFAX: 617/248-7100

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 6306 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..6306

PUBLICATION INFORMATION:

AUTHORS: COMPTON, DUANE A

AUTHORS: SZILAK, ILIYA

AUTHORS: CLEVELAND, DON W

TITLE: PRIMARY STRUCTURE OF OF NUMA, AN INTRANUCLEAR

TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR

TITLE: SEGREGATION OF PROTEINS AT MITOSIS

JOURNAL: JOURNAL OF CELL BIOLOGY

VOLUME: 116

PAGES: 1395-1408

DATE: MAR-1992

US-08-195-487-3

alignment\_scores:

Quality: 374.50

Ratio: 0.508

Percent Similarity: 49.596

Length: 1486

Gaps: 62

Percent Identity: 19.314

## Alignment block:

US-09-538-396-2 x US-08-195-487-3

Align seq 1/1 to: US-08-195-487-3 from: 1 to: 6306

```

103 GluLysAlaSerLysMetGluPheLysAlaIleGluSerValLeuGln.. 118
104 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
802 CAGCGCGCCAGCCACCTGAGAGCCAGAGAGCTTGAAGAGCTGGTGACAA 851
119 .....ThrlLeaProHisThrGlyLysValCys... 129
120 ||||| ||||| ||||| ||||| ||||| ||||| |||||
852 GAATGAGAGCTTACCATCGCGCTCATGAACCCCTGAAGCAGTGCAGG 901
130 ..LeuSerTyrArgCysAlaAspMetAspArgGluIleProAlaLeuMet 145
131 ||||| ||||| ||||| ||||| ||||| ||||| |||||
902 ACCGTGAAGACAGAGAGAGCCAGATGGATGCAAAATC..... 939
146 GlyValSerLysAlaValLeuGluLysValIlePheValHisGlnAspG 162
147 ||||| ||||| ||||| ||||| ||||| ||||| |||||
940 .....AACCACTTTCGAGAGAGAGATGAGAG 965
162 UserAsnTyrProLeuGlnAspProSerThrLeuLysLysLysPheAsp 179
163 ||||| ||||| ||||| ||||| ||||| ||||| |||||
966 CCTTCTTAAGCTGGGGAGTTGGCACTCATCTGCAGCAGCTACAGG 1015
179 sPrlPheSer....AlaThrArgTyrThrLysAla..... 189
180 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1016 ATGCCCTCAATGAGCTGACGAGAGAGACAGCAAGGCCACTCAGAGCTG 1065
190 LeuGluValIleLysLysLeuHisLysAsp.....GlnMetGlnI 203
191 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1066 CTAGAGAGAGAGCGCCAGCTGGAGAGAGAGCTACGCCACCTCGCAGGA 1115
203 uLysLysThrPheArgLysLysLeuGlnLysLeuGlnThrValLysAsp 220
204 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1116 CAAAGAAATGCTTGAGAGAGAGAAAGAAATCTCTCAG.....GAA 1156
220 LnaIAsnLysLysLeuArgLysAsnIleAlaGlnAspGlnGlnLysSerAsp 236
221 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1157 AACCTTCACAGCTGGAGAGAACACTTGTCCAGCTGCAGAGATAACCCACC 1206
237 AlasertLysSer.....GlnMetGlnLysLeu 246
238 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1207 CAGGAGAGAGCGCCAGCTGGTGGTGAATGCTTGCAGCTGCAACCTTGA 1256
246 sGluLysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSer 263
247 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1257 GCAGAGAGAGAGCCACTCTGTGTCGAACCAACACACAGCTCCAGCCAGG 1306
263 euAspGluLeuArgArgLysLeuGlnGlnIleAspIleLys...AlaThr 278
264 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1307 TAGAGATGCTGGAGAGCTGAGGAGCCAGAGAACCAAGCTGCTTGTCT 1356
279 GluArg.....SerThrLeuLeuThr 285
280 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1357 GAGCGGGGCCACTTTCGAAGAGAGAGAGAGAGAGCTGTCTAGCCGTATC 1406
285 TglnGlnHisGluLysLeuAlaLeuSerGluLysGlnLysAspThr 302
286 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1407 TGACCTGACAGAGCTCATCTCCAACTCAGAGAGAGAGAGAGAGAGCTGG 1456
302 sPrlGluLysLeuMetGluTyrPheGlnThrLysPheGluGluArgIleAla 318
303 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1457 AGCGAGGCTCCAGCTGATGAGAGAGAGAGAGAGAGAGCTGCTTGTCT 1506
319 LeuGluThrLysIleSerLysLeuValArgAspMetAspArgLys 335
320 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1507 CTGACCTCTGAGCTACACACTCAATGCCACCTCCAGCAGAGAGATCA 1556
335 rTyrSerSerValLeuSerLysGlnAsnSerGluLeuThrHisGluIleG 352
336 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1557 AGAACTGGCTGGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGCTAG 1606

```

```

352 LysLys...LeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHis... 366
353 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1607 CACAGACCTCTCCAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
367 .....GluArgAspSerAspIleLys 374
368 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1657 GTGAGCAGAGTAAGCACTAGCTGACAGAGAGAGAGAGAGAGAGAGAG 1706
374 nLleCysThrLysHisAsnLeuGlyProValProGlnHisProPheThr 391
375 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1707 GTTGGCGAGAGAG.....CAGAGAGCACTAGGC 1735
391 snAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLysSerSer 407
392 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1736 AGGACCATGCGCCAGCACTGGCCACTGCTCAGAGAGAGAGAGAGAGCTCC 1785
408 LeuGluAsn.....AspLeuLysAspLys 416
409 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1786 TTAAGGAGAGCGGATGCGCTCTCAGACAGCTGAGAGAGAGAGAGAG 1835
416 sLysSerAsnGlnAspGlnLeuAspValLeuThrPheLysHisTyrLeuLys 433
417 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1836 GAAGCT.....GCCAAGCTGGAGATTCTGACAGCAGCACTTCAGGTGG 1879
433 LeuAsn.....AlaArgTyrSerGluValAspGlyLys 443
434 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1880 CTATAGTAAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1929
444 .....IleGlnSerLysIleGlnSerMetSerGlyLeuArg 456
445 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1930 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1979
456 g.....ArgArgLysAspLysGluLysGluArgAspAlaIleGluValG 471
457 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1980 GACAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029
471 LuLeuSer...LysPheAsnLeuSerArgIleAspGluArgGluArgHis 486
472 ||||| ||||| ||||| ||||| ||||| ||||| |||||
2030 AGTTGACAGCTGGCTGTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 2079
487 MetGlnIleGluValGluArgLysThrLeuAlaLeuGluArgAspTyr 503
488 ||||| ||||| ||||| ||||| ||||| ||||| |||||
2080 GCCCAG.....GAGAGAGAG.. 2094
503 rAspSerIleLeuSerGlnLysArgThrGluValTyrSerLeuGlnLys 520
504 ||||| ||||| ||||| ||||| ||||| ||||| |||||
2095 .....CAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2128
520 yslLysValLeuLeuArgGluLysAspIleLeuAsnArgAsnAlaAsp 536
521 ||||| ||||| ||||| ||||| ||||| ||||| |||||
2129 CTTGAAAGGCTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2175
537 GluArgValLysLeuGlyLysLysAspAlaLeuGlnLysSerLysAs 553
538 ||||| ||||| ||||| ||||| ||||| ||||| |||||
2176 .....GCAGATGCCCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2201
553 PLYsLeuAsnGluIleValAsnGlnHisLysAspLysIleLysIleVal 570
554 ||||| ||||| ||||| ||||| ||||| ||||| |||||
2202 TTTGATCTCTGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2251
570 euArgGlyArgAsnProPheGlnLysAsp..... 579
571 ||||| ||||| ||||| ||||| ||||| ||||| |||||
2252 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2301
579 ..... 579
2302 GAGGCTGATTAAGCTGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 2351
580 MetLysLysGluIleAsnGlnAlaPhe.....TyrProValAspL 593
581 ||||| ||||| ||||| ||||| ||||| ||||| |||||
2352 CTTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2401

```

```

593 ysgIuTYrAsnGluLeu..... 598
594 .....
2402 GTGAGTGTGACAGCTCTCTCAAGAAAGTACGTGCGTGGAGCGGTAT 2451
599 ..ArgSerLysSerGlnGluAlaGluGlnLeuLysPheThrGlnSe 614
2452 GAGATACCCAGCAAGAGAGACACATGATGCGCCATGTTT...CAGGA 2498
614 rLyValThrAspAlaArgGluGlnLeuThrLysLeuArgArgAspMet 630
2499 ACAGCTGATGACTTTGAAGGAGGAATGTGAGAAAGCCCGCAGAGAGCTGC 2548
631 ..AspAlaLysArgArgPheLeuAspSerLysLeuGlnSerIleLeuGln 646
2549 AGGAGGCAAAAGAGAGAGTGGCAGGCTATAGATCCACAGCGAGCTCCAG 2598
647 IleSer.....AlaAsnValAspMet 653
2599 ATAAGCCCGGAGCAGCAACAACCTAGCTGAGCTCCATGCCAACCTG.... 2643
653 rPheProLysValLeuGlnAspAlaMetAsnLysArgAspGluGlnLysA 670
2644 ....GCCAGAGCACTCCAGAGCTCCAAAGAGAAAGTGAAGTCCAGGCCAGA 2689
670 rGluGlnAsnPheAlaAsnGlnMetArgGluMetLeuAlaPro..... 684
2690 AGCTTGAGATGACTCTCCACTCTGCAGAAAGATGAGTGCACACAGC 2739
685 .....PheGlnHisLeuAlaArgLys..... 691
2740 AAAGAGTGGCCCGCTTGAGACCTTGTCGCCAAGGCGAGTGGAGACGA 2789
692 .....AsnHisValLysProCysGluArgAlaPheThrP 704
2790 GGAACAGCCTCCCGGAGATTAGTCAAG.....GAGCCTGGAGGCGAG 2833
704 rAspGluGlnAspGluPheValLysGlnArg...MetGlnAsnSer 719
2834 GAGACAGACAGCCCGAGTGGTGGAGAGCAACAGGAGCCAGTCTGC 2883
720 SerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSerAsnAlaG 736
2884 ACCACACAGCAGCGCTGCAGGCTATGAGCGGGAG.....GCAGA 2924
736 uAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyAspAlaTyV 753
2925 GCAGATGGGCAATGAGCTGGAACGCGCTGCGG.....G 2956
753 aLlysLeuValGluGluThrIleProLeuAlaGluLysAsnLeuAsnGln 769
2957 CGCGCTGATGGAGAGCGGAGCAGCAGCAGAGAGAGCGTGGCGAGCAG 3006
770 .....HisLeuAlaAspGluSerGlnLysAlaGlnAla... 780
3007 GAAAGGAGAGTGGCGCGCTGACCCAGAGAGCGGCGCTGCCAGGCTGA 3056
780 ..... 780
3057 CCTTGCCTGGAGAGAGCGGCGCAGACAGACTTGAGATGCGGCTGCAGA 3106
781 .....PheAspAspLeuLeuGlyValLeu 788
3107 ACGCCCTCAACGAGCAGCGTGTGAGTTCCTTACCTCGCAAGAGGCACTG 3156
789 AlaHisValGlnMetAspArgAspAlaValGlnAlaLeuLeuGlnProTh 805
3157 GCTCATGCCCTGACGGAAGAGAGAGCAGCAGAGATGGGCCAAGCT 3206
805 rAspThrIleAsp.....ArgHisValHisGluIleGlnGlnL 818
3207 TCGTGTCTGGAGCAGACCCAGATAAAGAGCTGAGAACTTGGCGCAA 3256
818 euValLysGluVal.....GluAspLeuGluTyAla 828

```

```

3237 CCGTGAAGCAACTGAAGAGACAGCTGGCTTAAGAAAGAAAGAGACAGCA 3306
829 LeuAspSerSerGlyArgGlyValLysSer..... 838
3307 .....TCTGGCTCAGAGCCCAATCTAGAGCTGTGCGAGAGACGA 3347
839 .....LeuGluGluIleGlnLeuGlnLysAsnPheLeuG 850
3348 GCCAACAGAGCCCAAGCTGAAAGCACTGCGGGCGAGAGGTGAGCAAGCTG 3397
850 LnaArg..... 851
3398 AACAGCAATGCCAGAGACAGACAGAGCTGACAGCTTGAACAGCAGC 3447
852 .....ThrArgAspThrLeuIleValGln 859
3448 CTGAGAGCTGACGGGCTCCCGGGCTGAGCGGAGACAGTCTCTGGAGAC 3497
859 uValAsp...AspLeuArgAspGlnHisArgMetLeuAsnGluAspMetS 875
3498 TCTGCAGGCGCAGTTAAGAGAGAGCCAGAGCTAGGCGACAGTCAGA 3547
875 er.....SerAlaGln.....Val 879
3548 GTGCCTTACCTCGGCCCAAGCGAGATTGGCTCTCCGACCAAGATTA 3597
880 ArgThrHisAsnAlaArgGluGluLysValLysAlaSerSerIleLeuGln 896
3598 CAAGACCAAGCAAGCTGAAGATGAGTGAAGGCCAGTGGCCCGGGGG 3647
896 uArgPheGlnLysSerGluGluGluLeuValLeuLeuAlaGluGluLysG 913
3648 CCGGCAGAGGCTGAGAGAGAAATAGCTCATCAGACCTTGGAGAGAG 3697
913 GlnGluLeuValGluLysLysLeuLeuGluGluSerLeuAspProLeu 929
3698 AGCTGTCCATCCTGATCCCGAGCTCTGAGAAAGAGGGAG..... 3741
930 SerLysGluLysGluSerLeuLeuGlnGlnLysArgLysGlnLys 946
3742 AGCAAGAGATTGAAGCGGCTGTGATGCGGAGTCAAGAAAGCCAGAA 3791
946 sLeuAspGlu..... 949
3792 GCTGGAGAGAGAGTGGCGCTGTCAGAGCAGAGACCCAGCAACAGTGC 3841
949 ..... 949
3842 CAGAGCTGCAGAACGAGCTGTCTGTGGGAGAGAGTGCAGAGCCTCC 3891
950 .....GluTyThrHisGlnLeuAlaGlnArg...LysArgGln 960
3892 GGGAGGAGAGCTGAGAAACAGCGGTGGCTTCAAGAAACCTGCGGAGGA 3941
960 uPheGlnGlnGlnLeuAspAlaLeuGlyArgLeuAsnMetLysIleLysG 977
3942 GCTGACCTCAGAGCTGAGCGTGGAGAGAGCTGGGCGCAAGATTGAAG 3991
977 LysTyLeuAspSer.....LysLysAsnGluLysLeuLysGluLeuGln 991
3992 CGTGGCAGAGAGATTCTTCCAGAAAGAGAGCGCTTCCACCTGCAG 4041
992 GlyArgHisVal..... 995
4042 CTCGAGACACACAGCAGAGCCCTGTGTGATGATGAGCTGTGCAGCTAA 4091
996 .....LeuCysHisSerGlnLeuGlnSerCysMetAlaLysGlnLnaArgT 1011
4092 GCACCTTGGCAGACAGCTGAGGCGGAGCGGCGTCCGGAAGAAAGCC 4141
1011 leSerAlaGluLeuAsnLysSerLys..... 1019

```

4142 ACCGTGAGAGCTGAGACAGACAGAGCCGCTGGGGAGCTCGGGCA 4191  
 1020 GlnLeuLeuGlnGlnGlnGlnLeuLysArgAsnIleAspAsnLe 1036  
 4192 GAGCTGCTGGGGCCAGCGAGAGCTGGGGAGCTGATCTCTCGCGCA 4241  
 1036 UlystYrArgLysThrLysAlaAspValGlnGlnLeuThrArgAspIleG 1053  
 4242 GAGGTGGAGAGAGAGACCGAAGACGTCAGAGCTCGGGCAGAGAGG 4291  
 1053 IuSerLeuGlnGlnArgLeuLeuSerIleGlySerLeuSerAlaIleGlu 1069  
 4292 CCAGCTTATGAGACAGACCTGAGCTG..... 4320  
 1070 AlaAspLeuLysArgHisSerGlnGlnGlnArgLeuAsnSerGluPh 1086  
 4321 .....AAGAAGCGCATGGC.....CTGCTGGCAGAGGA 4349  
 1086 eAsnArgTrpGlnGlnThrLeuSerValTyrGlnSerAsnIleSerLysH 1103  
 4350 GAACTGGGGGCTGGGTAG.....CGGGCAACCTTGGCGGC 4387  
 1103 IAsYsGlnGlnLeuLysLeuSerGlnTyrLysAspIleGlnLysArgTyr 1119  
 4388 AGTTTCGGAAGTGGAGTTGGACCGCCGGGNA.....AAGTAT 4428  
 1120 ThrAsnGlnPheLeuGlnLeuLysThr.....ThrGlnMetAlaAs 1133  
 4429 GTCCAAAGGTGGCAGCGCTGCTGATGCTGAGACCGTCTGCTGA 4478  
 1133 nLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMet... 1148  
 4479 GGTCCAGGAGAGACACAGACACTGCCGGGAGCTGAGGTATGACTG 4528  
 1149 ..ArgPheHisSerMetLys.....MetGlnGlnLeuAsnLysIle 1161  
 4529 CCAAGTATGAGGTGCCAAGTCAAGGCTCGAGAGAGAGCAGCGTTC 4578  
 1162 IleLysGlnLeuTrpGlnGlnThrTyrArgGlnLysAspIleAspTyrI 1178  
 4579 CAGGAAAGAGAGAGAACTCATCTGCCAGGTGGAGAGAACTGAGTAAGA 4628  
 1178 eSerIleAsnSerAspSerGlnGlnArgLysThrArgSerTyrSerTyrA 1195  
 4629 ACTGCTGACTCTGACCAAGCCAGCAGAGTGCACAGCAGAACTGAAGG 4678  
 1195 rGValValMetGlnThrGlyAspAlaGlnLeuGlnMetArgGlyArgCys 1211  
 4679 CTGTCCAGGCTCAGAGAGGCGAGAGCCAGCAGAGGCCACGCTTCAG 4728  
 1212 SerAlaGlnGlnLysValLeuAlaSerLeuIleIleArgLeuAlaLeuAl 1228  
 4729 GCCCAGCTGATGAACTCAAGCCAGCTTGAAGCCAGAGAGAGCAGCAGC 4778  
 1228 aGluThrPheLysLeuAsnCysGlyIleLeuAlaLeuAspLupProHrT 1245  
 4779 TGAGCACTATTAAGCTGAG.....ATGGAGAAACCCAAA 4813  
 1245 hAsnLeuAspGlyProAsnAlaGlnSer.....LeuAlaAlaLeu 1259  
 4814 CACATATATATGCCAAGAGCAGAGAAACCAAGCTCAGAGACAGCTG 4863  
 1260 LeuArgIleMetGlnAlaArgLysGlnLysAsnPheGlnLeuIleVal 1276  
 4864 CGGAGCTGTGAGACAGCTGCAAG.....GAAAACAAAGAGCTG... 4902  
 1276 IleThrHisArgLysArgPheAlaHisLeuIleGlnArgGlnLeuAl 1293  
 4903 ..CGACTTAACCTGAACGGCTGGCCATGAGCTACAGAGGCTGGGCTGA 4951  
 1293 LaGlnLys 1295  
 4952 AGACCAAG 4959

seq\_name: /cgn2\_6/ptodata/2/lna/PCTUS\_COMB.seq:PCT-US93-06160-3  
 seq\_documentation\_block:  
 ; Sequence 3, Application PC/TUS9306160  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
 ; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: TESTA HURWITZ & THIBEAULT  
 ; STREET: 53 STATE STREET  
 ; CITY: BOSTON  
 ; STATE: MA USA  
 ; COUNTRY: USA  
 ; ZIP: 02109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/06160  
 ; FILING DATE: 19930621  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: PITCHER ESQ, EDMUND R  
 ; REGISTRATION NUMBER: 27,829  
 ; REFERENCE/DOCKET NUMBER: MTP-013  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617/248-7000  
 ; TELEFAX: 617/248-7100  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 6306 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..6306  
 ; PUBLICATION INFORMATION:  
 ; AUTHORS: COMPTON, DUANE A  
 ; AUTHORS: SZILAK, ILITA  
 ; TITLE: PRIMARY STRUCTURE OF OF NUMA, AN INTRANUCLEAR  
 ; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR  
 ; TITLE: SEGREGATION OF PROTEINS AT MITOSIS  
 ; JOURNAL: JOURNAL OF CELL BIOLOGY  
 ; VOLUME: 116  
 ; PAGES: 1395-1408  
 ; DATE: MAR-1992  
 ; PCT-US93-06160-3  
 alignment\_scores:  
 ; Quality: 374.50 Length: 1486  
 ; Ratio: 0.508 Gaps: 62  
 ; Percent Similarity: 49.596 Percent Identity: 19.314  
 alignment\_block:  
 US-09-538-396-2 x PCT-US93-06160-3 ..  
 Align seg 1/1 to: PCT-US93-06160-3 from: 1 to: 6306  
 103 GlnLysAlaSerLysMetGlnPheLysAlaIleGlnSerValLeuGln.. 118  
 802 CAGCGGGCCAGCCAGCTGAGCCCAAGAGAGCTTGAAGAGCTGCTGACAA 851  
 119 .....ThrIleAsnProHisThrGlyGlnLysValCys.... 129

852 GAATGAGACCTTACCATGGCGCTGCATGAAACCTGAAGCAGTGCAGG 901  
130 ..LeuSerTyrArgCysAlaAspMetAspArgGluIleProAlaLeuMet 145  
||| : : : : :  
902 ACCTGAAGACAGAGAAAGACAGATGATCCGAAATC..... 939  
146 GlyValSerLysAlaValLeuGluAsnValIlePheValHisGlnAspGlu 162  
||| : : : : :  
940 .....AACCAAGCTTTCGAGAGAGAAATGGAGAA 965  
152 userAsnTrpProLeuGlnAspProSerThrLeuLysLysLysPheAspA 179  
||| : : : : :  
966 CTTTCCTTTAAAGCTGCGGAGATTGCCAGTATCTGCAGCAGCTACAGG 1015  
179 spIlePheSer.....AlaThrArgTyrThrLysAla..... 189  
|| : : : : :  
1016 ATGCCCTCAATAGCTGACGGAGAGACAGACAGAGCCACTCAGAGAGTGG 1065  
190 LeuGluValIleLysLysLeuHisLysAsp.....GlnMetGlnGlu 203  
||| : : : : :  
1066 CTAGAGAACACAGCCCGCAGCTGAGAGAGAGCTCAGCGCAGCCCTGCAGGA 1115  
203 uIleLysThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspG 220  
||| : : : : :  
1116 CAGAAATGCTTGAAGAGAAAGAAATCCTTCAG.....GGA 1156  
220 InAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnLysSerAsp 236  
|| : : : : :  
1157 ACCTTTCACAGCTGGAGAAACACTTGTCCAGCTGCAGAGATAACCCACCC 1206  
237 AlasLysSer.....GlnMetGlnGluLeu 246  
||| : : : : :  
1207 CAGAGAAAGCGGAGGTGCTGGGTATGTTCTGCAGCTGAAACCTTGAA 1256  
246 sGluLysLysGlyThrGluArgGluIleLeuGlnMetGluThrSerL 263  
||| : : : : :  
1257 GCAGAGGCGAGCACTCTTGTGCAAAACACACACAGCTCCAAAGCCAGGG 1306  
263 euAspGluLeuArgArgLeuGlnGlyGlnIleAspIleLys.....AlaThr 278  
||| : : : : :  
1307 TAGAGATGCTGGAGACTGAGCGAGCCAGACAGAAACCAAGCTGCTG 1356  
279 GluArg.....SerThrLeuLeu 285  
||| : : : : :  
1357 GAGCGGCGCACTTCGAAAGAAAGACAGACAGCTGTACCTGATCAC 1406  
285 rGlnGlnHisGluLysLeuAlaAlaLeuSerGluGluAsnGluAspThra 302  
||| : : : : :  
1407 TGACCTGCAGAGCTCCATCTCCAACTCAGCCAGGCCCAAGAAAGAGCTGG 1456  
302 spGluGluLeuMetGluTrpGlnThrLysPheGluGluArgIleAlaLeu 318  
||| : : : : :  
1457 AGCAGGCTCCCGAGCTCATGGGCGCGGTGACTGCCAGTGGCTCT 1506  
319 LeuGluThrLysIleSerLysLeuValArgAspMetAspAspGluAla 335  
||| : : : : :  
1507 CTGACCTCGAGCTCACACACTCAATGCCACCATCCACAGACAGATCA 1556  
335 rTyrSerSerValLeuSerLysGlnAsnSerGluLeuThrHisGluLeu 352  
||| : : : : :  
1557 AGAACTGGCTGGCTGAAAGCAGAGCCCAAGAAAGACAGGCCCGAGTAG 1606  
352 Lys.....LeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHis... 366  
||| : : : : :  
1607 CACAGACCCCTCCAAACAGAAACAGAGCCCTCCAGGCTCCGCGACAG 1656  
367 .....GluArgAspSerAspIleLysAs 374  
1657 GTGAGACAGCTAAAGACTAGCTGAAAGACAGAGAGCAGCAGTTGAAGA 1706  
374 nIleCysThrLysHisAsnLeuGluProValProGlnHisProIleHis 391  
||| : : : : :  
1707 GTAGAGCGAGAG.....CAGAGAGCAACTAGGC 1735

391 snAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSerSer 407  
||| : : : : :  
1736 AGGACCAAGCCACAGCACTGGCCACTGCTGCAGAGAGAGAGGCCCTCC 1785  
408 LeuGluAsn.....AspLeuLeuAspLysLys 416  
||| : : : : :  
1786 TTAAAGGAGCGGAGATGGCTCTCAAGCAGCTGAGAGCAGCTGAAAGGA 1835  
416 LysSerSerAsnGluAspGlnLeuAspValLeuTrpLysHisTyrLeuLysI 433  
||| : : : : :  
1836 GAAGCT.....GCCAAGCTGAGATTCTGCAGAGCAACTTCAGGTGG 1879  
433 LeAsn.....AlaArgTyrSerGluValAspGluLys... 443  
||| : : : : :  
1880 CTAAATGAACCCCGGACACTGCCAGACTCTCAGTGCACAGGCCCAAGCG 1929  
444 .....IleGlnSerLysIleGlnSerMetSerGlyLeuLeu 456  
||| : : : : :  
1930 GAGAAAGCAGAGCTGACCGGAAAGGTGAGGAATCTCCAGGCTGTGTGA 1979  
456 g.....ArgArgLysAspLysGluLysGluArgAspAlaAlaGluValG 471  
||| : : : : :  
1980 GACAGCCCGCCAGGAACACAGATGAGGCCAGGCCCGCAGGTTCAGAGCTG 2029  
471 LuLeuSer...LysPheAsnLeuSerArgIleAspGluArgGluArgHis 486  
||| : : : : :  
2030 AGTGTGAGCTGGGTGCTGACACCAAAAGCACTGAGAAAGAAAGGCTG 2079  
487 MetGlnIleGluValGluArgLysThrLeuAlaLeuGluArgLysArgTyr 503  
||| : : : : :  
2080 GCCCAG.....GAGAAGAGC... 2094  
503 rasPserIleIleSerGlnLysArgThrGluValTyrSerLeuGluGlnL 520  
||| : : : : :  
2095 .....CACTCCAGAGAGAGCCTCCAGGCCCTCAAGAGCT 2128  
520 ysrIleLysValLeuLeuArgGluLysAspIleIleAsnArgAsnAlaAsp 536  
||| : : : : :  
2129 CTTGAGAGCTGACCAAGGCGAGCTTGAAGAGAGAGAGAGCGAGCT... 2175  
537 GluArgValLysLeuGluLysLysAspAlaLeuGlnSerSerLysAs 553  
||| : : : : :  
2176 .....GCAGATCCCTGGAAGAGCAGCAGCG 2201  
553 PLYLeuAsnGluIleValAsnGluHisLysAspLysIleLysLysValL 570  
||| : : : : :  
2202 TTGATCTCTGAGCTGAAGCAGAGACCCGAGCCTGTGTGAGACAGATA 2251  
570 euArgGluArgAsnProPheGluLysAsp..... 579  
||| : : : : :  
2252 AGCGGGAACGAAAGAGCTGGAAGAAAGAGGCTGGCGCAAGGCGCTG 2301  
579 ..... 579  
2302 GAGGCTGATTAAGCTGAGCTGGGAGGCCCATCAGGCTGAGACTGAGT 2351  
580 MetLysLysGluIleAsnGlnAlaPhe.....TrpProValAspL 593  
||| : : : : :  
2352 CCGCGCGCGGAGCTGGCAGAGCCATGCTGCCAGACACAGACTGAGA 2401  
593 ysrGluTyrAsnGluLeu..... 598  
||| : : : : :  
2402 GTGAGTGTGAGAGCTGCTCAAGAACTAGCTGGCTGGCTGAGCGGTAT 2451  
599 ...ArgSerLysSerGlnGluAlaGluGlnGluLeuLysPheThrGln 614  
||| : : : : :  
2452 GAGGATGCCAGCAAGAGAGAGCACTAAGGGCGCATGTTC...CAGGA 2498  
614 LysValThrAspAlaArgGluGlnLeuThrLysLeuArgArgMet... 630  
||| : : : : :  
2499 ACAGCTGATGACTTGAAGAGAGAAATGTGAAGAGGCCCGCCAGAGAGCTGC 2548



```

631 ..AspAlaIySArgArgPheLeuAspSerIySLeuGlnSerIleLeuGln 646
      ::::::::::::::::::::
2549 AGGAGGCAAAAGAGAGAGTGGCAGCAGCATAGATCCACAGCAGACCTCCAG 2598
      ::::::::::::::::::::
647 ILIser.....AlaAsnValAspMe 653
      ::::::::::::::::::::
2599 ATTAAGCCGGCAGCAGAACTAGCTGAGCTCCATGCCAACTG..... 2643
      ::::::::::::::::::::
653 tPheProIySValLeuGlnAspAlaMetAsnIySArgAspIyGlnIySArg 670
      ::::::::::::::::::::
2644 ...GCCAGACACTCCAGCAGCTCCAAAGAGAGAGTCCAGGGCCAGCA 2689
      ::::::::::::::::::::
670 tYLeuGlnAsnPheAlaAsnIyMetArgIyMetIleuAlaPro..... 684
      ::::::::::::::::::::
2690 AGCTTGACATGACCTCTCCACTCTGAGCAAAAGATGGCTGCCACAGC 2739
      ::::::::::::::::::::
685 .....PheGlnHisIleuAlaArgIyS..... 691
      ::::::::::::::::::::
2740 AAAGAGTGGCCCGCTGGAGACCTGGTCCGCAAGGCAGGTGACAGCA 2789
      ::::::::::::::::::::
692 .....AsnHisValIySProCySArgIyArgAlaPheThrP 704
      ::::::::::::::::::::
2790 GGAACACACCTCCCGGAGGTAGTCAAG.....GAGCCTGCGAGGGCAG 2833
      ::::::::::::::::::::
704 tOAspGluGlnAspIyPheValIySArgIySArg...MetGlnAsnSer 719
      ::::::::::::::::::::
2834 GAGACAGACAGCCGAGTGGTGGAAAGAGCAACAGGACGCCAGTTCTGC 2883
      ::::::::::::::::::::
720 SerThrAlaGluArgSerIyAlaIleuAlaMetGluSerSerAsnAlaGln 736
      ::::::::::::::::::::
2884 AGCACACAGCGCAGCTGACGCTATGAGCGGGAG.....GCAGA 2924
      ::::::::::::::::::::
736 uAlaLeuPheGlnGlnIleuAspIySLeuArgThrIleTyAspAlaTyv 753
      ::::::::::::::::::::
2925 GCAGATGGCAATGAGCTGGAAAGCGCTCCG.....G 2956
      ::::::::::::::::::::
753 aIySLeuValGlnIyThrIleProIeuaIyGlnIySAsnLeuGln 769
      ::::::::::::::::::::
2957 CCGGCTGTGAGAGAGCCAGGGCAGCAGCAGAGAGAGCCTGGCAGCAG 3006
      ::::::::::::::::::::
770 .....HisIleuAlaAspIySerGlnIySArgIyAlaGlnAla... 780
      ::::::::::::::::::::
3007 GAAAGGAGGTGGCGGCTGACCCAGAGCGGGCCGCTGCCAGGCTGA 3056
      ::::::::::::::::::::
780 ..... 780
      ::::::::::::::::::::
3057 CCTTGGCCCTGGAAAGGGCCAGCAGCAGCTGAGATGGCGCTGCAGA 3106
      ::::::::::::::::::::
781 .....PheAspAspIleuGlnIyValIleu 788
      ::::::::::::::::::::
3107 AGCCCTCAAGCAGCAGCGTGTGAGATTGCTACCTCCAGAGAGCAGCTG 3156
      ::::::::::::::::::::
789 AlaHisValGlnMetAspArgAspAlaValGlnAlaLeuGlnProth 805
      ::::::::::::::::::::
3157 GCTCAATGCCCTGACGGAAGAGCAAGCAGCAGAGTGTGCCAGCT 3206
      ::::::::::::::::::::
805 tAspThrIleAsp.....ArgHisValHisGlnIleGlnIle 818
      ::::::::::::::::::::
3207 TCTGGTCTGGAGGAGCCCAATATAAAGCTGAGAGAACTTGGCAAA 3256
      ::::::::::::::::::::
818 euValIySArgIyVal.....GluAspLeuGlnIyTyAla 828
      ::::::::::::::::::::
3257 CCGTGAAGCACTGAAGAGAAAGCTGTAAAGAAAGAAAGAGCAGCA 3306
      ::::::::::::::::::::
829 LeuAspSerSerGlyArgGlyValIySArg..... 838
      ::::::::::::::::::::
3307 .....TCTGGCTCAGAGAGCCCAATCTGAGGCTGTGCCAGCAGCA 3347
      ::::::::::::::::::::
839 .....LeuGlnIyIleGlnIleuGlnIleuAsnPheLeuG 850
      ::::::::::::::::::::
3348 GCCAAAGGCCCCAAGCTGAGAGCACTGCGGCGCAGGTGAGCAAGCTGG 3397
      ::::::::::::::::::::
850 InArg..... 851
      ::::::::::::::::::::

```

```

3398 AACAGCAATGCCAGAAACAGAGAGCAGGCTGACAGCTTGAAAGCAGC 3447
      ::::::::::::::::::::
852 .....ThrArgAspThrIleuIleValGln 859
      ::::::::::::::::::::
3448 CTCGAGGCTGAGCGGCTCCCGGCTGAGCGGAGCAGTGTCTGGAGAC 3497
      ::::::::::::::::::::
859 uValAsp...AspLeuArgAspGlnHisArgMetIleuAsnGlnAspMetS 875
      ::::::::::::::::::::
3498 TCTCCAGGGCCAGTTAGAGAGAAAGGCCCAAGAGCTTAGGGCAGCTCAG 3547
      ::::::::::::::::::::
875 er.....SerAlaGln.....Val 879
      ::::::::::::::::::::
3548 GTGGCTTAGCTCGGCCCAACGAGAGTGTGCTGCTCCGACCAAGATA 3597
      ::::::::::::::::::::
880 ArgThrIleAsnAlaArgGlnGlnIySValIySArgIySArgSerIleuGln 896
      ::::::::::::::::::::
3598 CAAGACCAACAGAGGCTGAAGATGAGAGAGGCCAGGTGGCCCGGGG 3647
      ::::::::::::::::::::
896 uArgPheGlnIySArgIyGlnIyLeuValIleuAlaGlnIyGlnIyS 913
      ::::::::::::::::::::
3648 CCGGCAGAGGCTGAGAGAGAAATAGCCTCATCAGCAGCTTGAGAGAG 3697
      ::::::::::::::::::::
913 IyGlnIleuIleValGlnIySArgIySLeuGlnIyGlnIySerLeuAspProIeu 929
      ::::::::::::::::::::
3698 AGGTCTCATCTCTGAATGCCAGAGCTCTGAGAAAGAGAGGGAG..... 3741
      ::::::::::::::::::::
930 SerIySArgIySArgIySerLeuGlnIyGlnIyThrAsnAlaIleuIySArg 946
      ::::::::::::::::::::
3742 AGCAAGAGATTGAAGCGGCTGTGATGCGCGAGTCAGAGAAAGCCAGAA 3791
      ::::::::::::::::::::
946 sLeuAspIyS..... 949
      ::::::::::::::::::::
3792 GCTGAGAGAGAGCTGCGCTCTGAGCAGCAGACAGCAGCAGCAGCTGC 3841
      ::::::::::::::::::::
949 ..... 949
      ::::::::::::::::::::
3842 CAGAGCTCAGAGACAGCAGCTCTCTGCGGAGAGAGGTGCAGAGCTCC 3891
      ::::::::::::::::::::
950 .....GluTyThrIleGlnIleuAlaGluArg.....LysArgGln 960
      ::::::::::::::::::::
3892 GGGAGGAGAGGTGAGAAACAGCGGGTGGCTTCAGAGAACTTCCGCGCAG 3941
      ::::::::::::::::::::
960 uPheGlnGlnIyLeuAspAlaIleuGlyArgIleuAsnMetIySArgIyS 977
      ::::::::::::::::::::
3942 GCTGACCTCAGAGGCTGAGCGTGGAGAGCTGGGCAAGATTTGAAG 3991
      ::::::::::::::::::::
977 tYThrIleuAspSer.....LysIySAsnGlnIySLeuIySArgIyGln 991
      ::::::::::::::::::::
3992 CGTGCGAGAGAGAGTTCTTCCAGAAAGAGCAGCCCTTCCACCTGCAG 4041
      ::::::::::::::::::::
992 GlyArgHisVal..... 995
      ::::::::::::::::::::
4042 CTCGAGACACACAGCAGACAGCGCTGTGAGTGAAGCTGCTCCAGCTAA 4091
      ::::::::::::::::::::
996 ...LeuCyHisIleSerGlnIleuGlnSerCySmetIalIySArgIyGlnArg 1011
      ::::::::::::::::::::
4092 GCACCTCTCCAGCAGCTGACGCGGAGAGAGCGCTGCGGAGAAAGGCC 4141
      ::::::::::::::::::::
1011 leSerAlaGlnIleuAsnIySArgIyS..... 1019
      ::::::::::::::::::::
4142 ACCGTGAGAGAGCTGAGAGAGCAGCAGCAGCCGCTGGGGGAGTGGCGGCA 4191
      ::::::::::::::::::::
1020 GluLeuIleuGlnIyGlnIyLeuIySArgAsnIleAspAspAsnIle 1036
      ::::::::::::::::::::
4192 GAGCTCTCGGGGCCAGCGGAGCTTGGGAGCTATTCCTCTCGCGGCA 4241
      ::::::::::::::::::::
1036 uIySArgIyGlnIyThrIyAlaAspValGlnIleuThrArgAspIleG 1053
      ::::::::::::::::::::
4242 GAAGGTGAGAGAGCAGAGCAGCAGCAGCTGAGCTGCGGCGCAGAGAG 4291
      ::::::::::::::::::::
1053 IySerIleuGlnIyArgIleuSerIleGlySerIleuSerAlaIleGln 1069
      ::::::::::::::::::::

```

4292 CCAGCTATGACAGACAGCTGAGCATGCTG..... 4320  
1070 AAlaspleuysargh1sserclnLylsglu1uargleuAensergluph 1086  
4321 .....AAGAAAGCCGATGCG.....CTGCTGGCAGAGGA 4349  
1086 easnArgrp1nLylsleuSerValTyrGlnSerAn1leSerLysH 1103  
4350 GAACCGGGGCTGGGTGAG.....CGGGCCAACTTGGCCGCGC 4387  
1103 1sLysglnLylsleuLysSerGlnTyrLysAsp1leGlnLysArgTyr 1119  
4388 AGTTCTGAGAGTGAGTGTGACACGAGCCGCGAA.....AAGTAT 4428  
1120 ThrAnsglnPheLynLylsleuLysThr.....ThrglMetAlas 1133  
4429 GTCCAGAGTGGGAGCCGCTGATGCTGATGCTGAGACCCGCTGCTGA 4478  
1133 nLysAspleuAspArgTyrTyrThra1leuAspLysAlaLeuMet... 1148  
4479 GGTGACGAGAGAGACAGACAGCTGCCCGGAGAGCTGGAGGTGATGACTG 4528  
1149 ..ArgPheH1sSerMetLys.....MetGlnGln1leAsnLys1le 1161  
4529 CCAAGTATGAGGGTCCCAAGGTCAAGTCTGTGAGAGAGAGCGCGGTTC 4578  
1162 1leLysglnLylsleuLysGlnThra1leuLysArgLysGlnAsp1leAspTyr 1178  
4579 CAG 4628  
1178 eSer1leAsnSerAspSerGlnGlnLylsleuLysArgLysSerTyr 1195  
4629 ACTGCTGCTGCTTGCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4678  
1195 rGValValMetGlnThra1leuLysAlaGlnLylsleuLysArgLys 1211  
4679 CTGTCCAGGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4728  
1212 SerAlaGlnLylsValLylsleuLysSerLylsleuLysArgLysLeu 1228  
4729 GCCCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4778  
1228 aGlnThra1leuLysSerGlnLylsleuLysArgLysSerTyr 1245  
4779 TGAGCACTATTAAGCTGAG.....ATGGAGAAAGCCAAA 4813  
1245 hrAsnLeuAspLysProAsnAlaGlnSer.....LeuAlaAlaLeu 1259  
4814 CACATATATATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4863  
1260 LeuArg1leMetGlnLylsleuLysGlnLysArgLysGlnAsp1le 1276  
4864 CGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4902  
1276 1leThra1leuLysSerGlnLylsleuLysArgLysSerTyr 1293  
4903 ..CGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4951  
1293 1aGlnLys 1295  
4952 AGACCAAG 4959  
seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-056-200-93  
seq\_documentation\_block:  
: Sequence 93, Application US/08056200  
: Patent No. 5615500  
: GENERAL INFORMATION:  
: APPLICANT: Steinert, Peter M.  
: APPLICANT: Lee, Seung-Chul  
: APPLICANT: Kim, In-Gyu  
: APPLICANT: Chung, Soo-Il  
: APPLICANT: Park, Sang-Chul

TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knodde, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-Apr-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054,001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-9502  
FAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-056-200-93  
alignment\_scores:  
Quality: 347.50 Length: 1138  
Ratio: 0.616 Gaps: 43  
Percent Similarity: 49.561 Percent Identity: 19.420  
alignment\_block:  
US-09-538-396-2 x US-08-056-200-93 ..  
Align seg 1/1 to: US-08-056-200-93 from: 1 to: 9551  
73 G1G1uThrGlnLysGlnLysGlnLysLysLeuArgPheLysThrAla1 89  
2824 GGAG 2873  
89 aG1LysAspValValLys1leArgSerPheGlnLysThrGlnLysAlas 106  
2874 GCGCGGAG 2919  
106 e1ysMetGlnPheLysAla1leGlnSerValLeuGlnThra1leuAsnPro 122  
2920 ..CGGAG 2967  
123 H1sThrLysGlnLysValLysLysSerTyrArgCysAlaAspMetAsp.. 138  
2968 CACGAAACTGAGAGAGTTT.....CCAGACGAGAGAGCA 2999

139 ..... ArgGluIleProAlaLeu.....M 145  
 3000 ACTGCGAAGCGCGGAGCTGTGAGCTGAGAGAGAGAGCGCCGAGAGAGA 3049  
 145 etGlyValSerLysAlaValLeuGluAsnValIlePheValHisGlnAsp 161  
 3050 AACGACGACAAAGCGCGAGCGCCAGACAGAGAGTGTCCAGAGAGAAAGA 3099  
 162 GluSerAsnTrpProLeuGlnAspProSerThrLeuLysLysLysPheAs 178  
 3100 GAGAAAGAGTGTGAGAGAGCGGAG...ACAGTGTCCGAAAGAAAGA 3146  
 178 PAspIle.....PheSerAlaThrArgTyrThrLysAlaLeuGluValI 193  
 3147 GAAATTGCGAGAGAGAGCGCCGACGCGCAAGAGAGAGTCCAGAGAGAGA 3196  
 193 LelyLysLysLeuHisLysAspGlnMetGlnGlnIleLysThrPheArgLeu 209  
 3197 AAGAGCGAGCTAGGAGAGCTGAGCGCGCAAGAGCTGAGAGAGAGCGCCAG 3246  
 210 LysLeuGluAsnLeuGlnThrValLysAspGlnAlaHisLysLeuArgG 226  
 3247 GAGGAAGAGCGACGACGACAAAGGCTGAGCGCGAGAGAGCAATAAGGCG 3296  
 226 uAsnIleAlaGlnAspGlnGluLysSerAspAlaSerLysSerGlnMetG 243  
 3297 CAAG.....CAGGAGGAGGAGAGCGCGGACGACAGAGAGAGAGCGCG 3340  
 243 LuginLeuLysGluLysLysLysGlyThrGluArgGluIleLeuGlnMet 259  
 3341 AGCGACGAGAGAGCGCGAGAGAGAGAGAGAGAGCGCGAGAGAGAGAGCTG 3390  
 260 GluTrpSerLeuAspGluLeuArgArgLeuGlnGlnIleAspIleLys 276  
 3391 AGCGCGGAG 3440  
 276 sAlaThrLysArgSerThrLeuLeuThrGlnGlnHisGluLysLeuAla 293  
 3441 GAGAGAGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3490  
 293 LalaSerGlnGluGluAsnGluAspThrAspGlnGluLeuMetGluTrpG 309  
 3491 GCGACGACGAGCTGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
 310 ThrLysPheGlnGluArgIleAlaLeuLeuGluThrLysIleSerLys 326  
 3541 CTGAGCGCGAGAGAGAGAGCTG..... 3561  
 326 uValArgAspMetAspAspGluAlaSerTyrSerSerValLeuSerLysG 343  
 3562 .....AGGC 3565  
 343 LAsnSerGluLeuThrHisGluIleGlyLysLeuGlnAlaGluAlaAsp 359  
 3566 GCGACGACGAGCTGAGCGCGAG...CAGACGCTGAGCGCGAGAGAGAGAG 3612  
 360 AlaHisLeuThrMetLysHisGluArgAspSerAspIleLysAsnIleCy 376  
 3613 CTGAGCGCGAGAGAGAGCTGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 3657  
 376 sThrLysHisAsnLeuGlyProValProGlnHisProPheThrAsnAsp 393  
 3657 ..... 3657  
 393 AlaIleMetAsnLeuThrAsnArgIleLysAlaArgLeuSerSerLeuGlu 409  
 3658 .....GAG 3660  
 410 AsnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLeuAspValLe 426  
 3661 CAGCAGAGCTAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3699  
 426 uTrpLysHisLysLysLysLysLysLysLysLysLysLysLysLysLysLys 443

3700 .....CAGCTGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3727  
 443 LniIleGlnSerLysIleGluSerMetSerGlyIleLeuArgArgArgLys 459  
 3728 GCGCAGCGAG 3777  
 460 AspLysGluLysGluArgAspAlaAlaGluValGluLeuSerLysPheAs 476  
 3778 GAGCAGAGAGAGAGAGCGCGAGTGTG..... 3801  
 476 nLeuSerArgIleAspGluArgGluArgHisMetGlnIleGluValGluA 493  
 3802 CTGAAAGCGGAG 3841  
 493 rGlyThrLeuAlaLeuGlnGluArgAspTyrAspSerIleIleSerGln 509  
 3842 GCGCAG 3879  
 510 LysArgThrGluValTyrSerLeuGlnGlnLysIleLysValLeuLeuAr 526  
 3880 AGCGCGAG 3918  
 526 gGluLysAspIleIleAsnArgAsnAlaAspGluArgValLysLeuGlyL 543  
 3919 .....CGCAG 3949  
 543 eLysLysAspAlaLeuGluSerSerLysAspLysLysAsnGluIleVal 559  
 3950 TAAAGCGGAG 3993  
 560 AsnGluHisLysAspLysIleLysLysValLeuArgGluArgValAsnProPh 576  
 3994 CAG 4043  
 576 eGluLysAspMetLysLysGluIleAsnGlnAlaPheTrpProValAspL 593  
 4044 AAGAGCGAG 4081  
 593 ysgLutYrAsnGluLeuArgSerLysSerGlnGluAlaGlnGlnGluLeu 609  
 4082 AGCGCGAG 4131  
 610 LysPheThrGlnSerLysValThrAspAlaArgGluGlnLeuThrLysLe 626  
 4132 AAGCGCGAG 4166  
 626 uArgArgAspMetAspAlaLysArgArgPheLeuAspSerLysLeuGlns 643  
 4167 GAGCGCGAG 4200  
 643 eTlleLeuGlnIleSerAlaAsnValAspMetPheProLysValLeuGln 659  
 4200 ..... 4200  
 660 AspAlaMetAsnLysArgAspGluGlnLysArgLeuGluAsnPheAlaAs 676  
 4201 .....AAGCGCGAG 4227  
 676 nGlyMetArgLysMetLeuAlaProPheGlnHisLeuAlaArgLysAsnH 693  
 4228 .....CAGGAGCTGAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4248  
 693 LsValCysProCysCysGluArgAlaPheThrProAspGluGluAspGlu 709  
 4249 .....CTGAG 4284  
 710 PheValLysLysGlnArgMetGlnAsnSerSerThrAlaGluArgSerLys 726  
 4285 CCGCTGAAGCGCGAG 4334  
 726 sAlaLeuAlaMetGluSerSerAspAlaGluAlaLeuPheGlnGlnLeuA 743

4335 GAGCAGAGCAGAGAGAGAGCCGCCAGCACTGAGCGCGAGCAGC 4384  
743 splysluauagtttletyraspalatyraltylsluvalglutThr 759  
4385 AGGAAAGCGCC.....GAGCAGCGCGCTGAAAGCCGAGAGAGAG... 4422  
760 lleroleuaglulysasnleuasnlnhisleuallaspglusergl 776  
4423 .....GAGAGAGGCTGACAGAGCGGCTGAAAGCCGAGCATGA 4460  
776 nlysalaglinalpheaspasplesleuglyvalleuallhisvalglim 793  
4461 GGAAGAGAGCGCGCAGCAGAGCTA..... 4485  
793 etaspaayaspalavalglualaleuenglnprothrAspThrileasp 809  
4486 .....GCTGAGAGAGAGCAGCAAGACGCCGCGGACGCTTAAG 4524  
810 ArgHisValHisgluilegln...GlnleuValLysgluValglu.... 823  
4525 AGCCGCATCCGAGAGTGCGACAGCTAGAAAGCGAGCGCAGCAGCAG 4574  
824 AspleuaglulTyrAlaleuaspSerSerGlyArgGlyValLysSerLeug 840  
4575 GCAAGCAAAAGTCTTACTCGAGCGCCCGCAAGCAGAGAGCGCAGCGCC 4624  
840 lugalileglnleuglnleuasnphleuglnArgThrArgAspThrleu 856  
4625 CGCAAGACCGCAGAGCA.....AAGAGCGCGCGCGCAGAGTGAG 4662  
857 lIleValgluValAspAspleuArgAspGlnHisArgmetleuasngluas 873  
4663 CTGCAATGCGAGAGAGAGAGAGCG...GCTCACCGCGCAGCAGCAGAGA 4709  
873 pmetSereralaglnValArgThrHisasnAlaArgglutLysValL 890  
4710 GGAGCAGCGCGCGGACTGCACATGCGCAGTGCAGCGCAGAGAAAGAGCG 4759  
890 ysAlaSerSerlleuengluarGpneGln.....LysSer 901  
4760 AGAGGGGCGCT.....CAGAGGCTGTCGCGCAGCGCCCATTCGCGGAG 4803  
902 GluGluGluLeuValLeuLeuaglulglulLysgluLnuLeuIleValgl 918  
4804 CAGCGGAGAGGAGCTGAGGCGCGAGGAGCGCCAGCAG..... 4842  
918 ulysLysleuenglulgluserleuaspProleuserLysgluLysglus 935  
4843 .....CGGAAACAAGGTTTCTCCGCGAGAGAGAGAGAGAGCAGC 4882  
935 erleuenglulglutyrasnAlaLeuLysgluLysleu..... 947  
4883 AGCGCGCGCGCTCCAGAGAGATCAGAGAGAGCGCAGAGCAGAGCGCC 4932  
948 AspGluGluTyrHisgluLeuaglulArgLysArgGluPheGlnGlnl 964  
4933 GAAGAGAGCAGCTCCAGCGCGGAGCGTCCGCAACAGCTCCAGAGAGA 4982  
964 uleuaspAlaleu.....GlyArgL 971  
4983 CGAGAGCGCGCTCCAGAGAGATCAGAGAGAGCGCAGAGCAGAGCGCC 5032  
971 euasnMetLysIleLysgluTyrLeuaspSerLysLysasngluLysleu 987  
5033 GCGACCAAAAATGAGTGCAACTAGAAAGAAAGAAAGAGAGAGCG... 5079  
988 LysgluLeuGlnLysArgHisValleu...CysHisSerGlnleuGlnse 1003  
5080 .....CGCCACAGCGCTGTACGCCAAGCAGCCCTACAGAGA 5114  
1003 rCysMetAlaLysGlnLysArgLysSerAlaLysleuasnLysSerLysg 1020  
5115 GCAGCTGAGAGAGAACAGCAGCTGTCGAGAGAG.....GAGAGAGAGG 5158

1020 luleuenglulgluLysGlnleuLysArgasnIleaspAspAsnleu 1036  
5159 AGCTACAGAGAGAGAGCGCGAAGAGAAAGCGCCCAAGAACAGAGAGA 5208  
1037 LysTyrArgLysThrLysAlaaspValgluLnuLeuThrArgAspIlegl 1053  
5209 CAATACCGC.....GAGAAAGCAGCTGCAGCAGAGAG... 5241  
1053 userleuenglulArgleuLeuSerIlegLysSerleuserAlaIlegl 1070  
5242 .....GAGAGCAGCTGTGAGA.....GAGC 5263  
1070 laaspLeuLysArgHisSerGlnLysgluValgluLeuasnSerGluPhe 1086  
5264 AACGGAGAAAAGAAAGAGCGCCAGAGCGGAGAAAG... 5298  
1087 AsnArgTyrGlnLysThrLeuSerValTyrGlnSerAsnIleSerLysH 1103  
5299 .....CAATA 5303  
1103 slysglnleuLysLysSerGlnTyrLysAspIleglulysArgTyrT 1120  
5304 TCGAAGAGATAGAGCTGCAGCAGAGAGAGAG..... 5337  
1120 hrasnGlnPheleuGlnleuLysThrThrGluMetAlaasnLysAspleu 1136  
5338 .....CAGCTGCTGGAGAGAGAACCGGAGAAAGAGAGCGCGAGCGG 5382  
1137 AspArgTyrTyrThrAlaleuaspLysAlaleuMetArgPheHisSerMe 1153  
5383 GAGAAATAATACCGGAGAGAGAGAGGTTGCGACGAGAGAGAACACACT 5432  
1153 LysMetGluGluLysLysIlelleLysgluLeuTyrGlnGlnThrT 1170  
5433 GCTGAGAGAGAGAACGGAGAAAGAGAGCGCCAGAGAGTGGAGAGCAGT 5482  
1170 yTArgLysLinsp 1174  
5483 ACCGCAAAAAGAGC 5496

Thu Nov 8 09:02:35 2001

us-09-538-396-2.rn1



1207 MetArgGlyArgCysSerAlaGlyGlnValLeuAlaSerLeuIleI 1223  
 152 ATGCGAGGGAGGTGTAGTGTGGCCAAAGGTCTTGATCATCATCAT 201  
 1223 GATGLeuAlaLeuAlaGlnThrPheCysLeuAsnCysGlyIleLeuAla 1240  
 202 ACATATGGGTGGTGGTGAACCTTTGTCTCAACTCGGGAATCTAGCAC 251  
 1240 euAspGluProThrThrAsnLeuAspGlyProAsnAlaGluSerLeuAla 1256  
 252 TGGATTAACCTCCACAAACCTAGACGCCCAACAGTGAAGTCTTGCT 301  
 1257 AlaAlaLeuLeuArgIleMetGluAlaArgGlyGlnGluAsnProheg 1273  
 302 GCGAGCTGTAGATTAATGAGGACAGCAAGGCAAGAACTTCA 351  
 1273 nleuIleValIleThrHisAspGluArgPheAlaHisLeuIleGlyGln 1290  
 352 ACTAATGTCATTACACAGCATGAGGATTTGCTCAATATATGGCCAAC 401  
 1290 rglInLeuAlaGluGlyGlyTyrTyrArgValSerIysAspGluAsnGlnHis 1306  
 402 GCCAGATGCGAAAGTATATATGTCATCTCAAGACGACATCAGCAC 451  
 1307 SerIleIleGluSerGlnGluIlePheAsp 1316  
 452 AGTATATTCACAGCTCAAGAAATTTGAT 481  
 seq\_name: gb\_est30:AU100679

seq\_documentation\_block: 706 bp mRNA EST 22-AUG-2000  
 LOCUS AU100679 Rice callus Oryza sativa subsp. Japonica cDNA clone C11847  
 DEFINITION AU100679 Rice callus Oryza sativa subsp. Japonica cDNA clone C11847  
 ACCESSION AU100679  
 VERSION AU100679.1 GI:9864929  
 KEYWORDS EST.  
 SOURCE Oryza sativa subsp. Japonica.  
 ORGANISM Oryza sativa subsp. Japonica.  
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Eupharidiales; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 REFERENCE Sasaki, T. and Yamamoto, K.  
 1 (bases 1 to 706)  
 TITLE Rice cDNA from callus (2000)  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program  
 2-1-2 Kannondai, Tsukuba  
 Ibaraki,  
 Japan 305  
 Tel: 0298-38-7441  
 Fax: 0298-38-7468  
 Email: tsasaki@abr.affrc.go.jp  
 PROJECT -RGP  
 C11847\_6Z.

FEATURES  
 source  
 1..706  
 /organism="Oryza sativa subsp. Japonica"  
 /strain="cultivar Nipponbare, sub-species Japonica"  
 /db\_xref="taxon:39947"  
 /clone="C11847"  
 /clone\_1lb="Rice callus"  
 /note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
 NotI; cDNA prepared from rice callus mRNAs by using  
 oligo(dt) as a primer and ligating to the SalI-NotI site  
 of pBluescript II SK+ phagemid."  
 BASE COUNT 262 a 114 c 175 g 152 t 3 others  
 ORIGIN  
 alignment\_scores:

Quality: 700.00 Length: 233  
 Ratio: 3.500 Gaps: 4  
 Percent Similarity: 85.837 Percent Identity: 64.378

alignment\_block:  
 US-09-538-396-2 x AU100679 ..  
 Align seg 1/1 to: AU100679 from: 1 to: 706

781 PheAspAspLeuGluGlyValLeuAlaHisValGlnMetAspArgAsp 797  
 17 TTGATGATATTTTCTAGTGTCTTCTGCTCAAGTTAAATATGCCAAGATG 66  
 797 laval.GluAlaLeuLeuGlnProThr.AspThrIleAspArgHisValH 813  
 67 CAGTGCAGGTATGCTACAGCCAGTGAACGATTAACGATTAAGCAGCTG 116  
 813 lsgIleGlnGlnLeuValGlyGlnValGlnAspLeuGlyTyrAlaLeu 829  
 117 AAGAAATCCAGCAGTTAGCACCCACCAAGTTGAAATCTTGATATTAAGCTT 166  
 830 AspSerSerGlyArgGlyValIysSerLeuGlnGluIleGlnLeuGlu 846  
 167 GATGTCGTGGCCAGGTGTTAATCTTGGACCAATCCAGTTGGAGTT 216  
 846 uAsnPheLeuGlnArgThrArgAspThrLeuIleValGluValAspAsp 863  
 217 GAACCTGTGCACAGACCAAGGACACATTAACATGAAGTGAATGATC 266  
 863 euArgAspGlnHisArgMetLeuAsnGluAspSerSerAlaGlnVal 879  
 267 TAAGGATCAACAACAAACACTTACCGATGCTGCTCAATATGCTCAGATG 316  
 880 ArgTyrHisAsnAlaArgGlnGluGlyValIysAlaSerSerIleLeuG 896  
 317 CGATGCGATGATATTAGGAGAGAGAGCTAAAGCTTCAGTCCAGTCA 366  
 896 uArgPheGlnIysSerGlnGluGluLeuValLeuAlaGlnGluIysG 913  
 367 CAATATTCAAAAAGCTGACAGAGATTTGGGACATCTACGAGAGAGAA 416  
 913 lglInLeuIleValGluGlyLysLeuLeuGlnGluSerLeuAspProLeu 929  
 417 AAAAAGTACATTAGAAAGAACATTTAGAGAGATCTTGGTGGCTG 466  
 930 SerIysGlnIysSerLeuLeuGlnGluTyrAsnAlaLeuIysGln 946  
 467 TCAAAAGAGAGAGAAAGCTTGTGCAAGACATGAACCTTGAAGGAAA 516  
 946 sleuAspGluGluTyrHisGlnLeuAlaGluArgLysArgGluPheGln 963  
 517 GCTTGATCAGAGATATCATCACTGCGCAGAAAGAAAGGAATTCACG 566  
 963 lngInLeuAspAlaLeuGlyArgLeuAsnMetLysIleGlyTyrLeu 979  
 567 AGGAATATGATGACCTGACACACACACATAGCGTATCAAAAGGTACCTG 616  
 980 AspSerLysLysAsnGluLysLeuLysGlnLeuGlnGlyArgHisVal 996  
 617 AATTCAAAAGAGGCGGAGAGCTTAATGAACCTACAGGAAAAA..... 658  
 996 uCysHisSerGlnLeuGlnSerCysMetAlaLysGlnGlnArg 1010  
 659 ...CATATCAGTTCAGTCTGATCTACAAAAA..... 697

seq\_name: gb\_est77:BE637160

seq\_documentation\_block: 609 bp mRNA EST 25-AUG-2000  
 LOCUS BE637160  
 DEFINITION WHE1806\_H09\_O18Z Secale cereale anther cDNA library Secale cereale  
 cDNA clone WHE1806\_H09\_O18, mRNA sequence.  
 ACCESSION BE637160  
 VERSION BE637160.1 GI:9920271

```

KEYWORDS      EST.
SOURCE
ORGANISM      Secale cereale
               iye.
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
               ; Triliceae; Secale.
               1 (bases 1 to 609)
REFERENCE      Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Fenton
AUTHORS        R.D., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R.,
               Miller,R., Rausch,C.J., Ross,K., Seaton,C.L. and Tong,J.C.
               The structure and function of the expressed portion of the wheat
               genomes - Anther cDNA library from iye
               Unpublished (2000)
JOURNAL        Contact: Olin Anderson
COMMENT         US Department of Agriculture, Agriculture Research Service, Pacific
               West Area, Western Regional Research Center
               800 Buchanan Street, Albany, CA 94710, USA
               Tel: 5105595773
               Fax: 5105595818
               Email: oanderson@wpr.usda.gov
               Sequence have been trimmed to remove vector sequence and low
               quality sequence with phred score less than 20
               Seq primer: StrataGene SK primer.
FEATURES
SOURCE
               Location/Qualifiers
               1..609
               /organism="Secale cereale"
               /cultivar="Blanco"
               /db_xref="taxon:4550"
               /clone="WHE1806.H09.O18"
               /clone_lib="Secale cereale anther cDNA library"
               /tissue_type="Anther"
               /dev_stage="Adult plant before anthesis"
               /lab_host="E. coli SOLR"
               /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
               site 1: EcoRI; Site 2: XhoI; Plants were grown in the
               greenhouse. Anthers were harvested and pooled from early
               meiosis to late meiosis. The tissue, total RNA, and
               poly(A) RNA were prepared (Butler, Ross and Gustafson) at
               University of Missouri, Columbia. A cDNA library was
               made, and the cDNA clones were in vivo excised to give
               phescript phagemids in the TV Close lab (Choi, Close,
               Fenton) at the University of California, Riverside.
               Plasmid DNA preparations and DNA sequencing were performed
               in the OD Anderson lab (all other authors)."
BASE COUNT    219 a      100 c      148 g      142 t
ORIGIN
alignment_scores:
               Quality: 689.00      Length: 200
               Ratio: 3.786      Gaps: 0
Percent Similarity: 91.000      Percent Identity: 68.500
alignment_block:
US-09-358-2 x BE637160 ..
Align seg 1/1 to: BE637160 from: 1 to: 609
872  GUAspmetserAgiInaLARGTTPRHsASaALARGlUGlUGlU 888
      8  GAGAGTCATTCAATGCTCAGATGCGGTGGCAAGCTTTRGGGAAGAAA 57
      888  sVALlYsAlAserSerIleLeuGluUArGpHeGlnLysSerGluGluGluL 905
      58  ACRAAGAGCCTCAAGTGTATTGCTGAAGTCAAAAAGCGCGAAGAAGAT 107
      905  euValleuLeuAlaGluGluLysGluGluLeuIleValGluLysLysLeu 921
      108  TGGTACATTTGCGAGAGAAAGAGACAATGCTCCAGATGACGAAGCAT 157
      922  LeuGlUGlUserLeuAspProLeuSerLysGluLysLysLysLeuGlU 938
      158  TTAGAGAAAGCTCTTGTTCATGCTCAAAAGAGAGAGAAAGCTTGTCGA 207

```

```

938  ng1utYrAsnAlaLeuLySgInLyLeuAspOluInLyTrHsInLeuA 955
208  AGAGTATTAAGCTTTGAAGGAAAGGTTTGATCAGGATGTGATGACGCTGG 257
955  laGluArLyArGluUpheGInGInLyLeuAspAlaLeuGlyArGLeu 971
258  CAGAAAGAAAGGAGGATTCAGCAAGAAATCAGATGACTTGGAAACCTT 307
972  AsnMetYsIleLyGlyTrLeuAspSerLySAsnGluLyLeuLy 988
308  AACACACGCATCAAGGCTATCTGATTCACAAAGAAAGTGAAGAGCTTA 357
988  sGluDeuGInGlyArGHisValLeuCyHisSerGInLeuGInSerCyam 1005
358  TGAATTCAGGAAAGCATATACCTTATCCCTATCTACGTATACAGAAATGTG 407
1005  eTaLySgInGInArGlyIleSerAlaGluLeuAsnLySserLySgIleuA 1021
408  AGGCAAGAAAGCAAGACATCGGGGTGACCTTGACAAAGCAACAGCTTA 457
1022  LeuGInLyGInGlyInLeuLyArGAsnIleAspAspAsnLeuLySTY 1038
458  TTACGAGACCAAGATCATTTGAAAAGAAATATGATGACAAATCTGACATA 507
1038  rArGlyThrLyAlaAspValGluGInLeuThrArGAspIleGlySerL 1055
508  CAGGAAACAAAGCGCTGAGTGCATCTTACTCATGATATGAGAGTAC 557
1055  euGluGluArGLeuLeuSerIleGlySerLeuSerAlaIleGluAlaAsp 1071
558  TTGAAGATTAATGTGCTTCCATCGGACATGCTCTATAGAAAGCTGAT 607
seq_name: gb_est74:BE449712

seq_documentation_block:
LOCUS      BE449712          486 bp      mRNA          26-JUL-2000
DEFINITION ESr361150 tomato root, plants pre-anthesis, Cornell University
LYCopersicon esculentum cDNA clone CLEYIK13 similar to strong
similarity to A. thaliana putative RAD50 DNA repair protein,
mRNA sequence.
VERSION    BE449712
KEYWORDS   BE449712.1  GI:9455215
SOURCE     EST
ORGANISM   tomato.
            Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE  1 (bases 1 to 486)
            van der Hoeven,R.S., Garvin,D., Matern,A.L., Holt,I.E., Liang,F.,
            Upton,J., Hansen,T., Craven,W.B., Bowman,C.L., Ahn,S., Renning,C.M.,
            Frazer,C.M., Martin,G.B., Giovannoni,J. and Tanksley,S.D.
            Generation of ESTs from tomato root tissue
            Unpublished (1999)
            Contact: David Frisch
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@clemson.edu
            5 prime sequence.
FEATURES             Location/Qualifiers
     source           1..486
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="CLEYIK13"
                     /clone_lib="tomato root, plants pre-anthesis, Cornell
                     University"
                     /tissue_type="root"

```



/dev\_stage="plants in pre-anthesis stage"  
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; supplier: Tanksley; tissue supplied by Dave Garvin  
 (USDA-ARS, Ithaca, NY 14850)."  
 BASE COUNT 152 a 99 c 125 g 110 t  
 ORIGIN

alignment\_scores:  
 Quality: 655.00 Length: 134  
 Ratio: 4.962 Gaps: 0  
 Percent Similarity: 98.507 Percent Identity: 90.299

alignment\_block:  
 US-09-538-396-2 x BE449712 ..

Align seg 1/1 to: BE449712 from: 1 to: 486

```

1 MetSerThrValAspLysMetLeuIleLysGlyLeuArgSerPheAspPr 17
|||||
84 ATGACCTGTATGATATGCTCTATTAAGGCGATCCGAGATTGACCC 133
|||||
17 CAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuIleValG 34
|||||
134 GGAGAGACAAAAAGCTTACACCTTTTCCGACCATTAATCTCATCTCG 183
|||||
34 LysProAsnGlyAlaGlyLysThrThrIleIleGluCysLeuLysLeuSer 50
|||||
184 GAGCTACAGCGCGCCGAAAAACTACTGTAATGAGTCTTGAAAGTTGCT 233
|||||
51 CysThrGlyLysLeuProPheAsnSerArgSerGlyHisThrPheValH 67
|||||
234 TGCACCTGGTATATGCCGCCAAATGTCCTGCTGCTACGCTTATCCA 283
|||||
67 AspProLysValAlaGlyLysThrGluThrLysGlyGlnIleLysLeuA 84
|||||
284 CGACCTTAGAGTGGCTGGAGAGACGAGACAAAGGCGCATTAAGCTGC 333
|||||
84 TGPheLysThrAlaAlaGlyLysAspValValCysIleArgSerPheGln 100
|||||
334 GGTTCAAAACAGACAGCAGGAAAGATGTGTGCTATGAGCTTTTTCAG 383
|||||
101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerValle 117
|||||
384 CTAACCTCGAAGGCGACAAAGATGGAATACAGCCATTGAGAGGCTTCT 433
|||||
117 uGlnThrIleAsnProHisThrGlyLysValCysLeuSerTyrArgC 134
|||||
434 ACAAGCTTAATTAATCTCTACACGCGAGAGAAAGTTGCTTAAGTTAGAT 483
|||||
134 ys 134
||
484 GT 485

```

seq\_name: gb\_est75:BE515984

seq\_documentation\_block:  
 BE515984 634 bp mRNA EST 08-AUG-2000  
 LOCUS  
 DEFINITION WHE0607\_E01.J012A wheat ABA-treated embryo cDNA library Trilicium  
 accession BE515984  
 version BE515984.1 GI:9740010  
 keywords  
 SOURCE  
 ORGANISM  
 Trilicium aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
 ; Triticeae; Trilicium.  
 1 (bases 1 to 634)  
 REFERENCE  
 AUTHORS Anderson,O.D., Chao,S., Han,P.S., Hsia,C.C., Johnson,R.R., Kang,Y.,  
 Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Tong,J.C., Verhey,  
 S.D., and Walker-Simmons,M.K.  
 TITLE The structure and function of the expressed portion of the wheat

JOURNAL  
 Unpublished (2000)  
 COMMENT  
 Contact: Olin Anderson  
 US Department of Agriculture, Agriculture Research Service, Pacific  
 West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 5105595773  
 Fax: 5105595818  
 Email: oanderson@pw.usda.gov

FEATURES  
 source  
 1. 634  
 /organism="Trilicium aestivum"  
 /cultivar="Brevor (soft, white, winter, common wheat)"  
 /db\_xref="taxon:4565"  
 /clone="WHE0607\_E01\_J01"  
 /clone\_11d="Wheat ABA-treated embryo cDNA library"  
 /tissue\_type="Seed embryo"  
 /dev\_stage="Mature dormant seeds"  
 /lab\_host="E. coli DH12S"  
 /note="Vector: pGAD10; Site 1: EcoRI; Site 2: XhoI;  
 Embryos were cut from mature, dormant seeds and imbedded in  
 25 microm ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7,  
 for 12 hr at 22 C. The tissue, total RNA, and poly(A)  
 RNA were prepared by Steven Verhey in M.K.  
 Walker-Simmons's lab (USDA-ARS, Washington State Univ.,  
 Pullman, Washington 99164-6420). A cDNA library was made  
 by Clontech using a combination of random and oligo dT  
 primers. Library was plated and archived by Russell  
 Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid  
 DNA preparations and DNA sequencing were performed in the  
 OD Anderson lab (all other authors)."

BASE COUNT 146 a 144 c 111 g 233 t  
 ORIGIN

alignment\_scores:  
 Quality: 630.00 Length: 203  
 Ratio: 3.600 Gaps: 0  
 Percent Similarity: 86.207 Percent Identity: 59.113

alignment\_block:  
 US-09-538-396-2 x BE515984/rev ..

Align seg 1/1 to reverse of: BE515984 from: 1 to: 634

```

502 AspTyrAspSerIleIleSerGlnLysArgThrGluValTyrSerLeuG 518
|||||
632 GACTATGATTGATTATATGTCAGAAAGCGCTCAGAGATATATACCTTGA 583
|||||
518 uGlnLysIleLysValLeuLeuArgGluLysAspIleIleAsnArgAsnA 535
|||||
582 TCACAAGATTAAGACACTTCACCGAGAGAAAGATTAACATGACCACTGAG 533
|||||
535 LAAspGluArgValLysLeuGlyLysLysAspAlaLeuLysLeuSer 551
|||||
532 CTGATGACAGAGAGTAAATTAAGCTCAAGAGAGATGAGTGAAGATGTC 483
|||||
552 LysAspLysLeuAsnGluIleValAsnGluHisLysAspLysIleLysLy 568
|||||
482 AAGAGAAAGACTTAAGATATATGATGACCTAAGATTAATTTAGAAAG 433
|||||
568 sValLeuArgGlyArgAsnProPheGlnLysAspMetLysLysGluIleA 585
|||||
432 TGTCTTAAGGAGAGGCTTCCTCATGAGAGAGATGTCAAGAGGAGATTA 383
|||||
585 sngLAlaPheThrProValAspLysGluTyrAsnGluLeuArgSerLys 601
|||||
382 CTCAGAGCTTGGGTCTGCTAGACTCAGATTAACATGATTGACTCAAAA 333
|||||
602 SerGlnGluAlaGluGlnLeuLysPheThrGlnSerLysValThrAs 618

```

```

|||||
332 TCTCAGGAGCGGAAACACACTGTGAATTTGGCACAAATGAAATTTGATCG 283
618 PALAARGLUGInLeuThrLysLeuArgArgAspMetAspAlaLysArgA 635
|||||
282 TGCTAAAGCCACTTGCAGAAAGCTCCAAAAGTTTGGATGCAAAAAAGAA 233
635 rghLeuAspSerLysLeuGlnSerLleLeuGlnLleSerAlaAsnVal 651
|||||
232 AGCAATCTGAACTCGAACTTCATTCATTCATTCATTCATTCATTCATTC 183
652 AspMetPheProLysValLeuGlnAspAlaMetAsnLysArgAspGluG 668
|||||
182 AACGCTTACCCCAAGATTCCTGAAGATGCCATGCGACGAGAGATTAACA 133
668 nLysArgLeuGlnAsnPheAlaAsnGlyMetArgGluMetLeuAlaPro 685
|||||
132 GACAAATATATTCACTATGCTAAGGAAATGCGCAAAATGTATGAACCTT 83
685 heGlnHisLeuAlaArgLysAsnHisValCysProCysCysGluArgAla 701
|||||
82 TTGAAAGAGTGCGCCGACAGCATCACAGTGTCTGTGTGTGTGTGTGTGT 33
702 PheThrPro 704
|||||
32 TTCACACCT 24
seq_name: gb_est30:AU122773

```

```

seq_documentation_block:
LOCUS AU122773 901 bp mRNA EST 23-Oct-2000
DEFINITION AU122773 MAMMAL Homo sapiens cDNA clone MAMMA1003095 5', mRNA
sequence.
ACCESSION AU122773 GI:10947489
VERSION AU122773.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 901)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.

```

```

JOURNAL HRI human cDNA project
COMMENT Unpublished (2000)
Contact: Takeo Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp

```

```

FEATURES
Source HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1..901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMA1003095"
/clone_lib="MAMMA1"
/tissue_type="mammary gland"
/note="Vector: pME18SFL3"

```

```

BASE COUNT 326 a 145 c 198 g 229 t 3 others
ORIGIN

```

```

alignment_scores:
Quality: 627.50 Length: 304
Ratio: 2.716 Gaps: 8
Percent Similarity: 75.987 Percent Identity: 44.408

```

```

alignment_block:
us-09-538-396-2 x AU122773
Align seg 1/1 to: AU122773 from: 1 to: 901

```

```

9 IleuGlyIleArgSerPhe.....AspProAspAsnLysAsnValI 23
|||||
5 ATTCTGGCGGTGGCGAGTTTGGAAATAGAGACAAAGTAAGCAAAATAT 54
23 eThrPhePheLysProLeuThrLeuIleValGlyProAsnLysIleGly 40
55 CACTTTCCTTCACGCCCTTACAAATTTTGGTGGACCCCAATGGCGGGA 104
40 yThrThrIleIleGluCysLeuLysLeuSerCysThrGlyGluLeuPro 56
|||||
109 AACAGCAGCATCTTGAATGCTCAAAATATTTTGTACTGCGAGATTCCCT 154
57 ProAsnSerArgSerGlyHisThrPheValHisAspProLysValAlaG 73
155 CCTGGAACCAAA...GGAATACATTTGTACACGATCCCAAGTTGTCTCA 201
73 yGluThrGluThrLysGlyGlnIleLysLeuArgPheLysThrAlaIle 90
202 AGAAACAGATGTGAGAGCCAGCATTCGTGCATTTGCTGATGTCAATG 251
90 LysAspValValCysIleArgSerPheGlnLeuThrGlnLysAlaSer 106
|||||
252 GAGAACTTATGCTGTGCAGAAAGATCTATGCTGTGTACTCAGAAAGCAAA 301
107 LysMetGluPheLysAlaIleGluSerValLeuGlnThrIleAsnProH 123
302 AAGACAGATTTTAAACTGTGGAAGAGCAT...ACTAGAACAAAGCA 348
123 sThrGlyLysValCysLeuSerThrArgCysAlaAspMetAspArgG 140
349 T...GGTGAAGAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 395
140 LuleProAlaLeuMetGlyValSerLysAlaValLeuGlnLysAlaIle 156
396 AATGATCAGTCTCTTGGGGTTTCCAAAGCTGTGCTAAATATGTCATT 445
157 PheValHisGlnAspGluSerAsnTrpProLeuGlnAspProSerThr 173
|||||
446 TTCTGTCATCAGAGAAATCTTAATTGCTTAAGTGAAGAAAGCCTT 495
173 uLysLysLysPheAspAspIlePheSerAlaThrArgIleLysAlaL 190
496 GAAGCAAAAGTTTGATGAGATTTTTCAGCAACAAGATACATTTAAAGCCT 545
190 eGluValIleLysLysLeuHisLysAspGlnMetClnLuleLysThr 206
546 TAGAAACACTTCGCGAGGTACGTCAGACACAAAGTGAAGTAAGTAAGAA 595
207 PheArgLeuLysLeuGlnAsnLeuGlnThrValLysAspGlnAlaHis 223
596 TATCAATGGAATTAATATCTGAAGCAATATTAAGGAAAGAAAGCTGTGA 645
223 sLeuArgLysAsnIleAlaGlnAspGlnGlnLysSerAspAlaSerLys 240
646 GATTCGATCAGATTAACAAGTAAGCAAGCCAGTTAAACATCTTCAAG. 694
240 eGlnMetGluGlnLeuLysGlnLysIleCysGlyThrGluArgGluIle 256
695 .....GAAAT 700
257 LeuGln...MetGluThrSerLeuAspGluLeuArgArgLeuGlnGly 272
701 GTCAATCCCTATGAGATGAACTGATTCATTAAGATGCTATAAGAG 750
272 nIleAspIleLysAlaThrGluArgSerThrLeuLeuThrGlnGlnHis 289
751 AATGACATAT.....CTCTCTAAATAAATGAAGAACTGTGACAAATG 790

```

```

289 1utylsleialaleuSERGIUGLUAsgLUASPThrsplugluieu 305
    ||  :: |||||:::  ::|||  ::|||:::
791 AA..ATPAAgCCTTGcATPAGCCCGAAAGAAAGCAATGAGAAAGTATA 837
    :::::||||:
306 MetGIurpGln 309
    :::::||||:
838 GTGAAGTGAAG 849

seq_name: gb_est69:BE013657

seq_documentation_block:
LOCUS      BE013657          531 bp      mRNA
DEFINITION 125081 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
ACCESSION  BE013657
VERSION     BE013657.1  GI:8274637
KEYWORDS   EST.
SOURCE      pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
REFERENCE  1 (bases 1 to 531)
            Fahrenkrug,S.C., Fieking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keel,J.W.
            Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
            Unpublished (2000)
JOURNAL    Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smitht@maamr.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
PCR PRIMERS
FORWARD: AGGAAACAGCATGACCAC
BACKWARD: GTTTCACAGTCACAGC
Plate: 47  row: J  column: 19
Seq primer: ATTTAGGTGACACTATAG.
            Location/Qualifiers
FEATURES             Location/Qualifiers
     source           1..531
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_id="MARC 1Pig"
                     /tissue_type="pooled"
                     /lab_host="DH108"
                     /note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos"
BASE COUNT      174 a          98 c          121 g          138 t
ORIGIN
alignment_scores:
    Quality:      599.50          Length:      176
    Ratio:        3.893           Gaps:        1
    Percent Similarity: 87.500    Percent Identity: 66.477

alignment_block:
US-09-538-396-2 x BE013657  ..

Align seg 1/1 to: BE013657 from: 1 to: 531

1117  LysAtgTtYThAsnGlnPhelengInleuLysTThrThGluMetAlaAs 1133
    :::::||||  :::::|||||  :::::|||||  :::::|||||
2  GAAAGTATAGGGAATGCTGATTGTTATGAGAACACACAGACTTGTA 51
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1133  nLysAspleuAspArgTtYThrThAlaLeuAspLysAlaLeuMetArgp 1150
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
52  CAAGGATCGAGACATTTATTATAGACCTTACGCCAAGGATPAATAAAT 101
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

1150 hehisermetlysmetglugliuleaslylletlelysgultuarp 1166
|||||
102 TTCACGATTCGAAATGAGAAATACAAATATTTCGTACCTCTGG 151
1167 glinglnthrtatrglyglnaspileaspyrileserlleasnseras 1183
|||||
152 CGAAGTACCTATTCGTGACACAGATATTGAATCATATGAAATTCGATCTGA 201
|||||
1183 p.....sergluglylaaelythrtatrgserlysettyra 1195
|||||
202 TGCCCGTGAATAATGTATCAGCTTCATTAAGACGCGGAATTTCACACTACC 251
|||||
1195 rgyvalymetglnthrglyaspalaagluenculmetatrglyatrgcys 1211
|||||
252 GAGTGTGATGCTTAAGGAGACACAGCCTTGATATGCGAGACGCTGTC 301
|||||
1212 serlaaglyglnlysvallleualaaserleulleleatrgleualaaleua 1228
|||||
302 AGTGTGAGACAGAGATATTAGCTCATCATCATTCGCTAGCCTAGCCTGGC 351
|||||
1228 agluthrtphrecysleuasncysgelylleleualaaleuaspgluptthrt 1245
|||||
352 TGAAATATTCGCTTAACTGTGACATTCGCTTACGCTTGATGACCAACA 401
|||||
1245 hrasmleuaspglytprosnalaglsertleualaalaaleuleuaatrg 1261
|||||
402 CAATATTTCACCGTGAATAATATTGATCTCTTCACATCTCTGCTGTGAG 451
|||||
1262 llemetglualaarglysglyglnguasnpheglnleullevalleth 1278
|||||
452 ATAAATTAAGAGTCGCTCACAGCAGCTAATTCACACTTCGTGTATTCAC 501
|||||
1278 rthasapgluargrphelaahisleulle 1287
|||||
502 TCATGATGAAGATTTTGTGAACCTCTTA 529
|||||
seq_name: gb_est41:AM032232

seq_documentation_block:
LOCUS AM032232 588 bp mRNA EST 15-SEP-1999
DEFINITION EST275686 tomato callus, TAMU Lycopersicon esculentum cDNA clone
CLEC35N16, mRNA sequence.
ACCESSION AM032232
VERSION AM032232.1 GI:5890988
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 588)
Alcala,J., Vrebilov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Romning
,C.M., Friser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@clemson.edu
5 prime sequence.
Location/Qualifiers
1..588
/organism="Lycopersicon esculentum"
/cultivar="TA96"
/db_xref="taxon:4081"
/cgene="CLEC35N16"
/clone_lib="tomato callus, TAMU"

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 12:58:30 ; Search time 5823.37 Seconds  
(Without alignments)  
11931.430 Million cell updates/sec

Title: US-09-538-396-1

Perfect score: 4492  
Sequence: 1 aatcgacacgagtgatcc.....atctgacatcaaaaaaaaa 4492

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 30

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: gb\_ba1:\*  
17: gb\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_stb1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_ro1:\*  
95: gb\_ro2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4492	100.0	4492	9 AX047235	AX047235 Sequence

ALIGNMENTS

RESULT 1  
AX047235

LOCUS AX047235 4492 bp DNA PAT 15-DEC-2000  
 DEFINITION Sequence 1 from Patent WO0068404.  
 ACCESSION AX047235  
 VERSION AX047235.1 GI:11876516  
 KEYWORDS  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 4492)  
 Mahajan, P.B. and Shi, J.  
 Maize rad50 orthologue and use of thereof  
 Patent: WO 0068404-A 1 16-NOV-2000.  
 PIONEER HI-BRED INTERNATIONAL, INC. (US)  
 Location/Qualifiers  
 1. 4492  
 /organism="Zea mays"  
 /db\_xref="taxon:4577"  
 292..4242  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC18962.1"  
 /db\_xref="GI:11876517"  
 /translation="MSTVDKMLIKIGRSFDPDNKNYTFEKKPLTLIVGNGAKTTII  
 ECLKSCTEGELPNRSRGTHPHDPRVAGETEKQIKIRTAAGKDVICISPOLT  
 OKSKMFEPAIESVLDTINPHGEKVLSCDMQREIPALMGVSKAVLENLYVHO  
 DESNMFLQPSLTLEKRFDI FSATRYTKLEVKLHKQMOEIKFKRLKLELYVK  
 DQHKLRKRENAQOEDSKSQMQLKKEICQTEKEIIOMETSLDELRLQOQIDIK  
 ATERSLLTQOHEKLAASENEDTDEEMENOTFEERLALLETISLKVMDDEA  
 SYSSTYSKNSLETHERIGLOAMADAHLMKIDNICKNHLGKVPPEHPPTND  
 VANTLNRIKARLSLENDLIDKKNSQEDLYMKHYIKINARVSEVDGQIOSKTES  
 MSGILNRKDKERDAEVELSKENLSRIDENRMOLEVEKTLALGBRDYSIIS  
 OKRTVEYSLEOKIKVLIREKDIINRNADREVKLGKDLAESKDLNVTNKHKI  
 KYLRGNPFPEKMKKEIKQAEWVPVDEKVELSKQOEAEQSLKFTQSTVAREDT  
 KLRDMDARKREFLSKQSLQISAVNDEPKYLOAMKRDQKLENFANAMREML  
 APREHARRKHVPCCEBAFTPEDEEPFKOMNSSTAREKALAMESNNAELFO  
 QDLKRTIYDAYKLYERTIPLAKRLNHLADESKAOFDOLLIVAHVODRBY  
 EALLQPTDIDRHHVEIQQLKVEDELETLADSSGKVSLEIQLBLNLOTRPTL  
 LIVEVDLROHRLNEDMSAQVWNNANREKVRKASSILRFKSEBELVLAEREQ  
 LIVKELLESLELPSLEKESLELEYNALOKLDEBYHOLEKREFOEDLALGRIN  
 MKIKGLDSKKNELKELOGRHYLCSLOSQAQOIRISAEINKSEKELLOGQOKR  
 NIDNINKYRTKADVOLFTRDIELEERLISLSLEADLRHSEKRLSENR  
 WGTSLYSYOSNISKHOKELKLSYKDIERTYNOFLQKTEMANDLDDYYVALDKA  
 LMGFHSKMEIKKIIKELMOQYRGQDIDYISINSDBAGTRISYRYVMYTGAE  
 LEHRGCSAQKYLASLIRLALAEFLCMLGLADEPTNLDPAESLAALAKRI  
 MEARKCOENQLIVITHDERFAHLIGROLAERYKRVSKDENOHSIIESQIIPD"

BASE COUNT 1537 a 831 c 1061 g 1063 t  
 ORIGIN

Query Match 100.0%; Score 4492; DB 9; Length 4492;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 4492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

241 TGGATCGAGTGCAGTCTAGAGGAGGACATTGGGGCTCTGGGGCAAGATGACACC 300  
 QY 301 GTTGAAGAATGCTGATCAAGGGGATTCGGACCTCGATCCGGAACAATAAGACGTCATC 360  
 Db 301 GTTGAAGAATGCTGATCAAGGGGATTCGGACCTCGATCCGGAACAATAAGACGTCATC 360  
 QY 361 acccttccaagcgcctcaaccctcatcgttggcccaagtgctggcaagaccagatc 420  
 Db 361 ACTTCTTCAAGCGCGTCAACCTCATCTGTGGCCCAACGCTGCTGGCAAGACAGATC 420  
 QY 421 atcgaatgctgaagcttcttccacccggagacgtcccccaactccgctctggccac 480  
 Db 421 ATCGAGTGCCTGAAGCTTCTTCCACCGGAGACTGCCCCCAACTCCGCTTGGCCAC 480  
 QY 481 accctgtccagcaccaccaagtgatcgtggagacggaacaaagacaatlaagt 540  
 Db 481 ACCTTGTCACACACCCCAAGTAGTACGTGGGACACGGAACAAAGACCAATTAAGTTG 540  
 QY 541 cgtttaagactgcagcagaagaagatgtgtgtcattccggtcctccaagttaccaa 600  
 Db 541 CGGTTTAAGACTGCAGAGGAAGATGTGTGTCATCCGTCCTTCAGCTTACCCAA 600  
 QY 601 aaggcatcaaatgagatgttaagcgaattgaagagctcctccaagactataatccac 660  
 Db 601 AAGGATCAAAAGATGGAGTTTAAGGCAATTGAAGCGTCTCCAGACTATAAATCCAC 660  
 QY 661 acagggaagaatctgcctcagctacagatgtgtgtacatgtgataagagattccgtcc 720  
 Db 661 ACAGGGAGAAAGCTGCTCAGCTACAGATGTGTGTACATGTGATAGAGATTCCGTCC 720  
 QY 721 ttaatggtgtttcgaagcgttactggaagatgtlatattgtttccaagaatgaatcc 780  
 Db 721 TTAATGSGTGTTCGAAGGCGGTACGTAGAGATGTATATTGTTCACCAAGATGATCC 780  
 QY 781 aattgacattgcaagaccgcgtcaacacctaagaagaattgcatgcatctctccgc 840  
 Db 781 AATTGGCATTGCGAGACCGGTCAACCTTAAGAAAGATTCGATGACATCTTCTGCGC 840  
 QY 841 accgcgtatcgaagaatcttgaagatcataaagaactccaagaatccaatgcaag 900  
 Db 841 ACAGCGTATCGAAGGCTCTTGAAGTATTAAGAACTTCAAGAGATCAAAATGCAAGAG 900  
 QY 901 atcaagacttlaagttlaaagcttgaagaacttcaagcttlaaagaaccaagcatalaag 960  
 Db 901 ATCAAGACTTTTAACTTAAGCTGAGACACTTCAAGCTGTAAGAAAGCAAGCAGATAG 960  
 QY 961 ctgcgtgaaatctgtcctaagatcaagaagaatgcagatgcctcaaaaatctcagatgag 1020  
 Db 961 CTGCGTGAATAATATTGCTCAAGATCAAGAAAGTCAAGATGCTTCAAAATCTCAAGATGAG 1020  
 QY 1021 caactgaagaagaatctgtgtgacagagagagaatccctgaacaatggaagaagtgtg 1080  
 Db 1021 CAACTGAAGGAAGATCTGTGTGACGAGAGAGAAATCTCTGCAAAATGGAAGAAAGTTTG 1080  
 QY 1081 gatgaactgaagaacttcaagagacaatltgacatcaaggaacaagaagagatgacatla 1140  
 Db 1081 GATGAAGTGAAGAACTTCAAGGAGCAAAATTGACATCAAGGCAACAGAGAGATGATCAATTA 1140  
 QY 1141 cttaagcagcagcagcagaagaatctgtgtgacatcttgaaggaatgtgaatgacagatgag 1200  
 Db 1141 CTTAAGCAGCAGCAGTGAAGAAAGCTTGTGCACTTCTGAGGAAATGGAAGATCGATGAG 1200  
 QY 1201 gaactaagatgtgcaacaacaatttgaagaagaagattgtgttaccagagaacaataatc 1260  
 Db 1201 GAAGTAAATGAGATGCAACAAATTTGAGAAAGATTTGCTTACTTGAAGAAACAAATATC 1260  
 QY 1261 agtaaaccttgaagagatltgaltgaltgaagacatctataagctccgttctgtccaaaca 1320  
 Db 1261 AGTAACCTTGAAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
 QY 1321 aattcgtaatcaacatgaaattggaagaatccagcagcagaagatgtgtcaccatgact 1380  
 Db 1321 AATTGTAATTACATGAAATTTGAAAGCTCCAGGCAAGAGATGATGATGATGATGATGATGAT 1380

QY	1381	atgaagcatgaacgagactcaacataaataatcgcactaaacatactctggcgc	1440
Db	1381	ATGAAGCATGAACGAGACTCAACATAAATAATATGCACTAAACATAATCTTGGCGCG	1440
QY	1441	gttcctgaacatccctttaagatgaatgtgtgcataagaccttaacaacagattaaagc	1500
Db	1441	GTTCTCTGAACATCCCTTTACGAATGATGTTGCTATGAACTTACAAACAGGATTAAGCG	1500
QY	1501	agacatcaagctcttgaaatgaattctgtgataaagaagaatccaaatgaagatcacgta	1560
Db	1501	AGACTATCAAGCTCTTGAGATGATGTTCTGATTAAGAAGAAATCCAAATGAAAGATCACTTA	1560
QY	1561	gacgtcttgtgaaacactacttaanaataaatgctgcactccgaagttgaatggtcag	1620
Db	1561	GATGTTTGTGTGAACACTATCTTAAATAATGATGCTGCTATCCGAAGTTGATGTTGAC	1620
QY	1621	atacaatctaagttgaatccatgctcaagcattttaagccygaagaagaataagaaagaa	1680
Db	1621	ATACAACTTAAGATTGAATTCCAATGTCAGGCATTTTAAGACGGAGAAAAGATTAAGAGAAA	1680
QY	1681	gaacggaatgcgcagaaagtgaggtcttcaaaatlaarctatcccgatccgattgaag	1740
Db	1681	GAACGGATGCTCCAGAGTGGAGCTTTCAAAATTTATATTCCTGATTCGATGAAGAG	1740
QY	1741	gagagacatatgcaaatctgaatgcgagagaaacacctgcgtcttgagaagaagactat	1800
Db	1741	GAGAGACATATGCAAAATTGAAGTCGAGAGAAACACTTCCTGTGAGAAAGAGCTAT	1800
QY	1801	gattcaattataagatcgaagaacgacaagaagatatagtttggaacagaanaataaagt	1860
Db	1801	GATTCAATTATAGTTCGAAACGAAAGATATAGTTTGAACAGAAATTAAGTGC	1860
QY	1861	ctttctcggggagaaagattataataatagaabaatgctgatagaagaagtaaaactcgggtt	1920
Db	1861	CTTCTCGGGAGAAAGATTATTAATTAATAAAGTCGATGAAGAAAGATTAATACTGGGTTTG	1920
QY	1921	aagaagatgcattctgaaagacgacaaggaacaaagctcaatgagatagttaaatgagcaat	1980
Db	1921	AAGAAGATGCATTGGAAGCAACAGCAAGCAACACTCAATAGATAGTTAATGAGCATAG	1980
QY	1981	gataaaatcaaaaaggctactgaaggggaggaatccctttgagaagagataatgaagaagag	2040
Db	1981	GATAAAATCAAAAAGCTACTTGAAGGGAGGAAGATCCTTTGAGAGAGATATGAAGAGAG	2040
QY	2041	atcaatcaagccttttggcctgttgacaaggaatacaatgaagttaagatcaaaaatccca	2100
Db	2041	ATCAATCAAGCCTTTTGGCTGTGCAAGGAATTAATAAGATTAAGATCAAAATCCAC	2100
QY	2101	gaagcaagacaagagctttaaattactgaagcaaaagtactgaatctagaagaacaattg	2160
Db	2101	GAAAGCAAGACAAGACTTAAATTACTCAGAGCAAAAGTAACTGATGCTTAGAGACAAATTG	2160
QY	2161	acaaaactcgaaagagataatgatacgaaanaaagaatccctggaaccggaactcaact	2220
Db	2161	ACAAAACCTCGAAGAGATGATGATGCACAAAGAAGATTCCTGAGCTCGAACTTCATATC	2220
QY	2221	atttaacgatactcgtctaattgtatgaatgcttcccaagttctacaagccgcatgaac	2280
Db	2221	ATTTTACGATATCTGCTAATGTTGACATGTTCCCAAGTTCTACMAAGCCCATGAAC	2280
QY	2281	aaaagagatgaacgaaagaatctagaagatcttcgcaaatggaatgtgggaaatgtctga	2340
Db	2281	AAAAGAGATGAACAGAAAAGATTAGGAATTTCCGAATAGGAATGGGAAATGCTTTGCA	2340
QY	2341	cccttgaacattcgtctcgagaagaatcagtatgccatgctgtgaaagctcttcacaa	2400
Db	2341	CCTTTGAACATTTTGGCTCGGAAGATCATGTATGCCCATGCTGTGAAGCTGCTTCACA	2400
QY	2401	ccgtgtggggagatgagatctgtgaagaacaaagaagatgcaaaactcaagtactcgag	2460
Db	2401	CTGTATGGAGAGATGATGTTCTGTGAAGAAACAAAGATGCAAAACATCAAGTACTCGAGAG	2460

QY	2461	agatctaaagctctgggcaatgaaatcatcaaatgctggaagctcttttcaagcaattgcat	2520
QY	2461	aaattctaaagctctgggcaatgaaatcatcaaatgctggaagctcttttcaagcaattgcat	2520
QY	2521	aaacttcgacatcatatgactatgactatgaaactgtaagaagaacacatctcttagca	2580
QY	2521	aaacttcgacatcatatgactatgactatgaaactgtaagaagaacacatctcttagca	2580
QY	2581	gagaaaaacttgaatcaacatttggcgatgaagaagtcagaagcgcgagacatttgaat	2640
QY	2581	gagaaaaacttgaatcaacatttggcgatgaagaagtcagaagcgcgagacatttgaat	2640
QY	2641	ctttgggtgtcttgcgccatgttcaaaatggacaggaatgcagcttgaagccattcaaa	2700
QY	2641	ctttgggtgtcttgcgccatgttcaaaatggacaggaatgcagcttgaagccattcaaa	2700
QY	2701	cccatgtatatttggcagagcatgtatcaagaattcaacagctgtgaagaagtgaag	2760
QY	2701	cccatgtatatttggcagagcatgtatcaagaattcaacagctgtgaagaagtgaag	2760
QY	2761	gattcttgaatatgacatttgaattctcaatggtgcgaggtgtcaactcttttgaagaattcaa	2820
QY	2761	gattcttgaatatgacatttgaattctcaatggtgcgaggtgtcaactcttttgaagaattcaa	2820
QY	2821	ctggagctgaaacttctctgcagagaaacaagggacacattgattgttcgaagtgtgatctt	2880
QY	2821	ctggagctgaaacttctctgcagagaaacaagggacacattgattgttcgaagtgtgatctt	2880
QY	2881	agagatcaactatgaatgactcaaatgaagatatgtcaagtgtcaggtgaagtgtgacacat	2940
QY	2881	agagatcaactatgaatgactcaaatgaagatatgtcaagtgtcaggtgaagtgtgacacat	2940
QY	2941	gctcggagagagaaagatgaagaagctcttcaacatatgtgaagaatttcaaaaatctgaaag	3000
QY	2941	gctcggagagagaaagatgaagaagctcttcaacatatgtgaagaatttcaaaaatctgaaag	3000
QY	3001	gaattgtgtctctagctctgagaaagaacaactgattgttgaagaagaagcttcttgaag	3060
QY	3001	gaattgtgtctctagctctgagaaagaacaactgattgttgaagaagaagcttcttgaag	3060
QY	3061	gagctcttgaatccattgttccaaagagaagaagagcttgttgcagaagtataatgcttgg	3120
QY	3061	gagctcttgaatccattgttccaaagagaagaagagcttgttgcagaagtataatgcttgg	3120
QY	3121	aagcaaaagctcggatgaagaagatactcaagcttgcagaaagaagaagaagaggtcttcagcaa	3180
QY	3121	aagcaaaagctcggatgaagaagatactcaagcttgcagaaagaagaagaagaggtcttcagcaa	3180
QY	3181	gaacttgaatgctcttgaagaacttaatatgaagataaaaggttacttgaattccaaagaa	3240
QY	3181	gaacttgaatgctcttgaagaacttaatatgaagataaaaggttacttgaattccaaagaa	3240
QY	3241	aaagaaaaagcttgaagaattgcagaggaagagcatgttcttgcacatctcagttacagagt	3300
QY	3241	aaagaaaaagcttgaagaattgcagaggaagagcatgttcttgcacatctcagttacagagt	3300
QY	3301	tgcattgcacaaacagcagaagaatatcaagcttgggtttaaacaagagagaagaagacatcttcag	3360
QY	3301	tgcattgcacaaacagcagaagaatatcaagcttgggtttaaacaagagagaagaagacatcttcag	3360
QY	3361	ggccagagccagtttgaagaagaacatcttgaacaactcacaagtaacagaagaagaagct	3420
QY	3361	ggccagagccagtttgaagaagaacatcttgaacaactcacaagtaacagaagaagaagct	3420
QY	3421	gatttggaaacaaacttactctgtgatataagatcatttgaagaagaagcgtcttcaatgagt	3480
QY	3421	gatttggaaacaaacttactctgtgatataagatcatttgaagaagaagcgtcttcaatgagt	3480
QY	3481	agcttgcctctctatagaagctatctgaaagccatttcaagaagaagaagaagagcttaat	3540
QY	3481	agcttgcctctctatagaagctatctgaaagccatttcaagaagaagaagaagagcttaat	3540
QY	3541	tcagaatttacaagctgtgcagaggaacaacttctgttatacaagtaataatttcaagcac	3600



Gencore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 22:20:31 ; Search time 260.71 Seconds  
(without alignments)  
10818.665 Million cell updates/sec

Title: US-09-538-396-1

Perfect score: 4492

Sequence: 1 aattcgacagatgagatcc.....atctagcatcaaaaaaaaa 4492

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size: 30

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_0601:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	4492	100.0	4492 22 AAC66915 Maize RAD50 coding

# ALIGNMENTS

RESULT 1  
AAC66915  
ID AAC66915 standard; cDNA; 4492 BP.  
XX  
AC AAC66915;

XX	27-MAR-2001 (first entry)
DT	Maize RAD50 coding sequence.
XX	
DE	Maize RAD50 coding sequence.
XX	
KW	Maize; RAD50; meiotic recombination; DNA repair; chromosome 4p; recombination efficiency; transformation efficiency; ss.
XX	
OS	Zea mays.
XX	
PN	MO200068404-A1.
XX	
PD	16-NOV-2000.
XX	
PF	25-APR-2000; 2000MO-US11086.
XX	
PR	05-MAY-1999; 99US-0132575.
XX	
PA	(PION-) PIONEER HI-BRED INT INC.
XX	
PI	MahaJan PB, Sh1 J;
XX	
DR	WPI: 2001-007402/01.
XX	
DR	P-PSDB: AAB27248.
XX	
PT	Nucleic acids encoding Rad50 polypeptides, useful for transgenic plant production and for producing antibodies that are useful for modifying gene expression -
PT	
XX	
PS	Claim 1; Page 65-70; 76pp; English.
XX	
CC	The present sequence is the coding sequence of the maize RAD50 gene, found on maize chromosome 4p. RAD50 is involved in meiotic recombination and DNA repair, and the gene and its protein can be used to produce transgenic plants whose expression of the gene may be regulated. This can be useful in the regulation of transformation and recombination efficiency in plants.
CC	
XX	
SQ	Sequence 4492 BP; 1537 A; 831 C; 1061 G; 1063 T; 0 other;
Query Match 100.0%; Score 4492; DB 22; Length 4492;	
Best Local Similarity 100.0%; Pred. No. 0;	
Matches 4492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 aattcgacagatgagatccatagaccacatagccgtacacaaaccctaagaaccctaac 60
DB	1 aattcgacagatgagatccatagaccacatagccgtacacaaaccctaagaaccctaac 60
QY	61 cgtacacaaaccctaagaaccctagacgctgaagactccacaaacacgcatcttctc 120
DB	61 cgtacacaaaccctaagaaccctagacgctgaagactccacaaacacgcatcttctc 120
QY	121 ctccacgccccctcttctcttcccaatcggttgcaatcactagaagcgtatgaat 180
DB	121 ctccacgccccctcttctcttcccaatcggttgcaatcactagaagcgtatgaat 180
QY	181 agaagttataggagatagatccgaatctaggtttggggcaatcgcttggccagac 240
DB	181 agaagttataggagatagatccgaatctaggtttggggcaatcgcttggccagac 240
QY	241 tgaatcgagatgacagtcgttagaaggagacattggggctcgttggggaagatagacac 300
DB	241 tgaatcgagatgacagtcgttagaaggagacattggggctcgttggggaagatagacac 300
QY	301 gttgacaaagatgctgataaaggagattcgagcttcggacgacataaagaatcattc 360
DB	301 gttgacaaagatgctgataaaggagattcgagcttcggacgacataaagaatcattc 360
QY	361 acctcttcaagcgcctacacccatcggtggcccaacggtgctggcaaacacacgac 420
DB	361 acctcttcaagcgcctacacccatcggtggcccaacggtgctggcaaacacacgac 420



421 atcgatgctgaaagcttcttctgcaacgagagctgcccccaactccgctctgagcaac 480  
421 atcgatgctgaaagcttcttctgcaacgagagctgcccccaactccgctctgagcaac 480  
481 accctcgtcccaagaccccaaggttagctgagcagcaagcaacaagaagcaaatgaatg 540  
481 accctcgtcccaagaccccaaggttagctgagcagcaagcaacaagaagcaaatgaatg 540  
541 cggtttaagactcgcagcagcaagcaagatgctgctgcatccggtctctccagcttccaa 600  
541 cggtttaagactcgcagcagcaagcaagatgctgctgcatccggtctctccagcttccaa 600  
601 aagcgcatacaagatgaggtttaaggcaatctgaaagctctcccaactataatccacac 660  
601 aagcgcatacaagatgaggtttaaggcaatctgaaagctctcccaactataatccacac 660  
661 aagcgcatacaagatgaggtttaaggcaatctgaaagctctcccaactataatccacac 720  
661 aagcgcatacaagatgaggtttaaggcaatctgaaagctctcccaactataatccacac 720  
721 ttaatggtgttctcgaaagcgctactgagcaatgtatataattgttcccaagatgaatcc 780  
721 ttaatggtgttctcgaaagcgctactgagcaatgtatataattgttcccaagatgaatcc 780  
781 aattggtgttctcgaaagcgctactgagcaatgtatataattgttcccaagatgaatcc 840  
781 aattggtgttctcgaaagcgctactgagcaatgtatataattgttcccaagatgaatcc 840  
841 aagcgcatacaagatgaggtttaaggcaatctgaaagctctcccaactataatccacac 900  
841 aagcgcatacaagatgaggtttaaggcaatctgaaagctctcccaactataatccacac 900  
901 atcaaaactctttaggtttaagagctgagcaactctcagactgtaaaagccaaagcaataag 960  
901 atcaaaactctttaggtttaagagctgagcaactctcagactgtaaaagccaaagcaataag 960  
961 ctgctgtaaaatcttctcagaaatcagaaagctcagactgtaaaagccaaagcaataag 1020  
961 ctgctgtaaaatcttctcagaaatcagaaagctcagactgtaaaagccaaagcaataag 1020  
1021 caactgaaagaaagatctgctgctgtaacgagagcaatccctgcaaatctgaaagcaatg 1080  
1021 caactgaaagaaagatctgctgctgtaacgagagcaatccctgcaaatctgaaagcaatg 1080  
1081 gatagaactgaaagactctcagagcaaatctgacatacaagcaagagagatataatca 1140  
1081 gatagaactgaaagactctcagagcaaatctgacatacaagcaagagagatataatca 1140  
1141 cttaacgagcagcagatgaaagctgctgcaactctcagagcaaatctgaaagcaatgag 1200  
1141 cttaacgagcagcagatgaaagctgctgcaactctcagagcaaatctgaaagcaatgag 1200  
1201 gaaactaagaaagagcaagaaatctgaaagaaagatctgctactaagaaacaaatcc 1260  
1201 gaaactaagaaagagcaagaaatctgaaagaaagatctgctactaagaaacaaatcc 1260  
1261 agcaaaactgtaagagatgagatgagatgagatgagatgagatgagatgagatgagatg 1320  
1261 agcaaaactgtaagagatgagatgagatgagatgagatgagatgagatgagatgagatg 1320  
1321 aattctgaatcaacaatgaaatctgaaagctcagagcaagagctgagctgagctgagct 1380  
1321 aattctgaatcaacaatgaaatctgaaagctcagagcaagagctgagctgagctgagctg 1380  
1381 atgaagagatgaaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1440  
1381 atgaagagatgaaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1440  
1441 gtctcgaacacaccccttaacgaatgagatgagatgagatgagatgagatgagatgagatg 1500  
1441 gtctcgaacacaccccttaacgaatgagatgagatgagatgagatgagatgagatgagatg 1500  
1501 agactacaaagctctgagatgagatgagatgagatgagatgagatgagatgagatgagatg 1560  
1501 agactacaaagctctgagatgagatgagatgagatgagatgagatgagatgagatgagatg 1560

1501 agactacaaagctctgagatgagatgagatgagatgagatgagatgagatgagatgagatg 1560  
1501 agactacaaagctctgagatgagatgagatgagatgagatgagatgagatgagatgagatg 1560  
1561 gatgtcttgtaaaacactatctttaaataatgctgctactccgaagttagtgctag 1620  
1561 gatgtcttgtaaaacactatctttaaataatgctgctactccgaagttagtgctag 1620  
1621 atacaatcaagatgagatgagatgagatgagatgagatgagatgagatgagatgagatg 1680  
1621 atacaatcaagatgagatgagatgagatgagatgagatgagatgagatgagatgagatg 1680  
1681 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1740  
1681 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1740  
1741 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1800  
1741 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1800  
1801 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1860  
1801 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1860  
1861 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1920  
1861 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1920  
1921 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1980  
1921 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 1980  
1981 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2040  
1981 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2040  
2041 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2100  
2041 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2100  
2101 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2160  
2101 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2160  
2161 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2220  
2161 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2220  
2221 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2280  
2221 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2280  
2281 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2340  
2281 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2340  
2341 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2400  
2341 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2400  
2401 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2460  
2401 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2460  
2461 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2520  
2461 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2520  
2521 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2580  
2521 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2580  
2581 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2640  
2581 gaaacgagatgctgcaagagctgagatgagatgagatgagatgagatgagatgagatgagatg 2640

Dh 2581 gagaaaaacttgatcaacatttggcggatgaagtcaagaagcgcgagcatttgat 2640  
Qy 2641 ctttgggtgtcttgcccatgttcaaatggaaggaagcggttgaaacctattaca 2700  
Db 2641 ctttgggtgtcttgcccatgttcaaatggaaggaagcggttgaaacctattaca 2700  
Qy 2701 cccactgtactacttgacagcgcatgtacatgaatccaacagctaacccaagaataga 2760  
Db 2701 cccactgtactacttgacagcgcatgtactgaatccaacagctaacccaagaataga 2760  
Qy 2761 gattcttgaaatgacatttgatctctgtgcccgaagtgatcaagcttttgaggaatcca 2820  
Db 2761 gattcttgaaatgacatttgatctctgtgcccgaagtgatcaagcttttgaggaatcca 2820  
Qy 2821 ctggagcggaactttcttcgaaggaacaaaggaacacattgtctgcaagtgtgatctt 2880  
Db 2821 ctggagcggaactttcttcgaaggaacaaaggaacacattgtctgcaagtgtgatctt 2880  
Qy 2881 agagatcaacatagatcttaaatgaatgaatcatgtcaagtgtcaggttgatgagcaat 2940  
Db 2881 agagatcaacatagatcttaaatgaatgaatcatgtcaagtgtcaggttgatgagcaat 2940  
Qy 2941 gctcggagagaagaagtgaaagcttctagcatattggaagaatcccaaaatctgaag 3000  
Db 2941 gctcggagagaagaagtgaaagcttctagcatattggaagaatcccaaaatctgaag 3000  
Qy 3001 gaattgggtgtcttctgaggaagaaagaacacatgattgtgaagaagaagctttaga 3060  
Db 3001 gaattgggtgtcttctgaggaagaaagaacacatgattgtgaagaagaagctttaga 3060  
Qy 3061 gaattgggtgtcttctgaggaagaaagaacacatgattgtgaagaagaagctttaga 3120  
Db 3061 gaattgggtgtcttctgaggaagaaagaacacatgattgtgaagaagaagctttaga 3120  
Qy 3121 aagcaaaagcttgatgaagatcatcatcagcttgcaagaagaagaaggtgtccagca 3180  
Db 3121 aagcaaaagcttgatgaagatcatcatcagcttgcaagaagaagaaggtgtccagca 3180  
Qy 3181 gaactgtgtctcttggaagacttaatttgaaataaaagggtacttgatccagaaga 3240  
Db 3181 gaactgtgtctcttggaagacttaatttgaaataaaagggtacttgatccagaaga 3240  
Qy 3241 aacgaagaagcttaagaagatgcaggaagcgctgtcttgcattccagttacaagat 3300  
Db 3241 aacgaagaagcttaagaagatgcaggaagcgctgtcttgcattccagttacaagat 3300  
Qy 3301 tgcattgcaaaaacagcaaaagatcatcagcttgatgaacaaagagcaaaactactgca 3360  
Db 3301 tgcattgcaaaaacagcaaaagatcatcagcttgatgaacaaagagcaaaactactgca 3360  
Qy 3361 ggccagagcgcatggaagaagaacatgatacaatcccaagttacaggaagaagct 3420  
Db 3361 ggccagagcgcatggaagaagaacatgatacaatcccaagttacaggaagaagct 3420  
Qy 3421 gatctggaacacatctactcgcgataatgaatccacttgaaagaagcgcttccaatagt 3480  
Db 3421 gatctggaacacatctactcgcgataatgaatccacttgaaagaagcgcttccaatagt 3480  
Qy 3481 agcttgctgtatagaagcgatctgaaacgcatctccaagaaagaagaggttaatt 3540  
Db 3481 agcttgctgtatagaagcgatctgaaacgcatctccaagaaagaagaggttaatt 3540  
Qy 3541 tcagaatttaacagatggaagaagaacattctgttttaacaaagtaattccaagac 3600  
Db 3541 tcagaatttaacagatggaagaagaacattctgttttaacaaagtaattccaagac 3600  
Qy 3601 aaacaagaagcttaacatctcaacatagaagatctcgaagaagcgatataactaacttc 3660  
Db 3601 aaacaagaagcttaacatctcaacatagaagatctcgaagaagcgatataactaacttc 3660  
Qy 3661 ctccagcttaagacaactgaatggaagaagaagacttggaagataatatactgcttta 3720  
Db 3661 ctccagcttaagacaactgaatggaagaagaagacttggaagataatatactgcttta 3720

Qy 3721 gacaagaagctcttatgtcggtttccacagcatgaagaatggaaggaataataataatcaag 3780  
Db 3721 gacaagaagctcttatgtcggtttccacagcatgaagaatggaaggaataataataatcaag 3780  
Qy 3781 gaactgtgcaacacagacatacagaagcgaggaattgatacctaagaacttaatttgat 3840  
Db 3781 gaactgtgcaacacagacatacagaagcgaggaattgatacctaagaacttaatttgat 3840  
Qy 3841 tctgaaggtgtctgcaatcgaatcatacagaactcagcgtgtgtatgcaaaactgtgtgct 3900  
Db 3841 tctgaaggtgtctgcaatcgaatcatacagaactcagcgtgtgtatgcaaaactgtgtgct 3900  
Qy 3901 gaggctggaatgcaagagcgctgcaagctgtctgtgcaagaaggtcttcttataatc 3960  
Db 3901 gaggctggaatgcaagagcgctgcaagctgtctgtgcaagaaggtcttcttcttataatc 3960  
Qy 3961 agactgacactgcaagaacttctgcgcgaactgaggaatttgatgattgataagca 4020  
Db 3961 agactgacactgcaagaacttctgcgcgaactgaggaatttgatgattgataagca 4020  
Qy 4021 actacgaatctagatgagcgaatgcaagaagctctgtcgtcgtcgtgtggaataatg 4080  
Db 4021 actacgaatctagatgagcgaatgcaagaagctctgtcgtcgtcgtgtggaataatg 4080  
Qy 4081 gaagcgaagaagaagcgaggaagaactccaagttgattgaatcaactcaatgaagaattt 4140  
Db 4081 gaagcgaagaagaagcgaggaagaactccaagttgattgaatcaactcaatgaagaattt 4140  
Qy 4141 gcccatcttatgctgcaagaagcgagcttggttgaaaggtactcaggtctccaagaatag 4200  
Db 4141 gcccatcttatgctgcaagaagcgagcttggttgaaaggtactcaggtctccaagaatag 4200  
Qy 4201 aaccagcaacagataattgaatcccaagaagaattgaactgaaggtgttctgaagagctg 4260  
Db 4201 aaccagcaacagataattgaatcccaagaagaattgaactgaaggtgttctgaagagctg 4260  
Qy 4261 tagcaagcactggttctgtatgcaatccagtttaattatgccaagtaactgtgccaag 4320  
Db 4261 tagcaagcactggttctgtatgcaatccagtttaattatgccaagtaactgtgccaag 4320  
Qy 4321 caattttaaagaagctttaggaagcctgtgttaagccggttaagcgaattgctgaagaactat 4380  
Db 4321 caattttaaagaagcctttaggaagcctgtgttaagccggtttagcgaattgctgaagaactat 4380  
Qy 4381 ccttggtgtgtatatactatttaactctgccaagaagatggaatgtgcaactggtgat 4440  
Db 4381 ccttggtgtgtatatactatttaactctgccaagaagatggaatgtgcaactggtgat 4440  
Qy 4441 ggaatttcaacagacatgaatgaatgttcaacaatcgaatcaaaaaaa 4492  
Db 4441 ggaatttcaacagacatgaatgaatgttcaacaatcgaatcaaaaaaa 4492

Search completed: November 8, 2001, 00:59:38  
Job time: 9547 sec

THU NOV 8 09:02:32 2001

us-09-538-396-1.rng

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 12:33:50 ; Search time 3157.09 Seconds  
(without alignments)  
13449.796 Million cell updates/sec

Title: US-09-538-396-1  
Perfect score: 4492  
Sequence: 1 aattggcgcagtgatcc.....atctagcatcaaaaaaaaa 4492

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 30

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

117: gb\_est48:\*  
 118: gb\_est49:\*  
 119: gb\_est50:\*  
 120: gb\_est51:\*  
 121: gb\_est52:\*  
 122: gb\_est53:\*  
 123: gb\_est54:\*  
 124: gb\_est55:\*  
 125: gb\_est56:\*  
 126: gb\_est57:\*  
 127: gb\_est58:\*  
 128: gb\_est59:\*  
 129: gb\_est60:\*  
 130: gb\_est61:\*  
 131: gb\_est62:\*  
 132: gb\_est63:\*  
 133: gb\_est64:\*  
 134: gb\_est65:\*  
 135: gb\_est66:\*  
 136: gb\_est67:\*  
 137: gb\_est68:\*  
 138: gb\_est69:\*  
 139: gb\_est70:\*  
 140: gb\_est71:\*  
 141: gb\_est72:\*  
 142: gb\_est73:\*  
 143: gb\_est74:\*  
 144: gb\_est75:\*  
 145: gb\_est76:\*  
 146: gb\_est77:\*  
 147: gb\_est78:\*  
 148: gb\_est79:\*  
 149: gb\_est80:\*  
 150: gb\_est81:\*  
 151: gb\_est82:\*  
 152: gb\_est83:\*  
 153: gb\_est84:\*  
 154: gb\_est85:\*  
 155: gb\_est86:\*  
 156: gb\_est87:\*  
 157: gb\_est88:\*  
 158: gb\_est89:\*  
 159: gb\_est90:\*  
 160: gb\_est91:\*  
 161: gb\_est92:\*  
 162: gb\_est93:\*  
 163: gb\_est94:\*  
 164: gb\_est95:\*  
 165: gb\_est96:\*  
 166: gb\_est97:\*  
 167: gb\_est98:\*  
 168: gb\_est99:\*  
 169: gb\_est100:\*  
 170: gb\_est101:\*  
 171: gb\_est102:\*  
 172: gb\_est103:\*  
 173: gb\_est104:\*  
 174: gb\_est105:\*  
 175: gb\_est106:\*  
 176: gb\_est107:\*  
 177: gb\_est108:\*  
 178: gb\_est109:\*  
 179: gb\_est110:\*  
 180: gb\_est111:\*  
 181: gb\_est112:\*  
 182: gb\_est113:\*  
 183: gb\_est114:\*  
 184: gb\_est115:\*  
 185: gb\_est116:\*  
 186: gb\_est117:\*  
 187: gb\_est118:\*  
 188: gb\_est119:\*  
 189: gb\_est120:\*

190: gb\_est121:\*  
 191: gb\_est122:\*  
 192: gb\_est123:\*  
 193: gb\_est124:\*  
 194: gb\_est125:\*  
 195: gb\_est126:\*  
 196: gb\_est127:\*  
 197: gb\_est128:\*  
 198: gb\_est129:\*  
 199: gb\_est130:\*  
 200: gb\_est131:\*  
 201: gb\_est132:\*  
 202: gb\_est133:\*  
 203: gb\_est134:\*  
 204: gb\_est135:\*  
 205: gb\_est136:\*  
 206: gb\_est137:\*  
 207: gb\_est138:\*  
 208: gb\_est139:\*  
 209: gb\_est140:\*  
 210: gb\_est141:\*  
 211: gb\_est142:\*  
 212: gb\_est143:\*  
 213: gb\_est144:\*  
 214: gb\_est145:\*  
 215: gb\_est146:\*  
 216: gb\_est147:\*  
 217: gb\_est148:\*  
 218: gb\_est149:\*  
 219: gb\_est150:\*  
 220: gb\_est151:\*  
 221: gb\_est152:\*  
 222: gb\_est153:\*  
 223: gb\_est154:\*  
 224: gb\_est155:\*  
 225: gb\_est156:\*  
 226: gb\_est157:\*  
 227: gb\_est158:\*  
 228: gb\_est159:\*  
 229: gb\_est160:\*  
 230: gb\_est161:\*  
 231: gb\_est162:\*  
 232: gb\_est163:\*  
 233: gb\_est164:\*  
 234: gb\_est165:\*  
 235: gb\_est166:\*  
 236: gb\_est167:\*  
 237: gb\_est168:\*  
 238: gb\_est169:\*  
 239: gb\_est170:\*  
 240: gb\_est171:\*  
 241: gb\_est172:\*  
 242: gb\_est173:\*  
 243: gb\_est174:\*  
 244: gb\_est175:\*  
 245: gb\_est176:\*  
 246: gb\_est177:\*  
 247: gb\_est178:\*  
 248: gb\_est179:\*  
 249: gb\_est180:\*  
 250: gb\_est181:\*  
 251: gb\_est182:\*  
 252: gb\_est183:\*  
 253: gb\_est184:\*  
 254: gb\_est185:\*  
 255: gb\_est186:\*  
 256: gb\_est187:\*  
 257: gb\_est188:\*  
 258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



Thu Nov 8 09:02:33 2001

us-09-538-396-1.rst





392 GCCCAACGCTGTCGCAAGACACATCATCGATGCTGACCTTCT 441  
51 CysThrGluLeuProPheAsnSerArgSerGlyHisThrPheValHis 67  
442 TGACACGGGAGCTGCCCCCACTCCGCTGCGCACACCTTCGTCA 491  
67 AsnProLysValAlaGluGluThrGluThrGlyGlnIleLysLeu 84  
492 CGACCCCAAGGTACTGCGGAGCGAAGAAACAAAGACAAATTAGTTGC 541  
84 rPheLysThrAlaAlaGlyLysAspValValCysIleArgSerPheGln 100  
542 GGTTHAGACTGCAAGAGAAAGATGGTGTGTATCCGCTCTTCAG 591  
101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerVal 117  
592 CTTCACCAAAAGCATCAAAAGATGAGATTAAAGCAATTGAACGCTCT 641  
117 uGlnThrIleAsnProHisThrGlyGluValCysLeuSerIleArg 134  
642 CCAGACTTAATCCACACACAGGGGAGAAAGTCTGCTCAGCTACAGAT 691  
134 ySAlaAspMetAspArgGluIleProAlaLeuMetGlyValSerLysAla 150  
692 GTGCGACATGATAGAGAGATTCCTGCTTAATGGTGTTCAGAGCC 741  
151 ValLeuGluAsnValIlePheValHisGlnAspGluSerAsnTrpPro 167  
742 GTACGGAGAAATGTTATTTGTTCCACCAAGATGATCCAAATGGCAT 791  
167 uGlnAspProSerThrLeuLysLysLysPheAspAspIlePheSerAla 184  
792 GCAGAGACCCGTCACACTTAAGAAAGATGATGATCTCTCTGCA 841  
184 hrArgTyrThrLysAlaLeuGluValIleLysLysLeuHisLysAspGln 200  
842 CACGCTTAACGAAAGCTTTGATGATCAAAAGAACTTCACAGAGATCA 891  
201 MetGlnGluIleLysThrPheArgLeuLysLeuGluAsnLeuGlnThr 217  
892 ATGCAAGAGATCAAGACTTTAGGTTAAAGCTGAGAACCTTCAGACT 941  
217 lLysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnGlu 234  
942 AAAGACCAAGACATTAAGCTGCGGAAATATTGCTCAAGATCAAGAA 991  
234 ySerAspAlaSerLysSerGlnMetGluGlnLeuLysGluLysIleCys 250  
992 AGTCAGATGCTCAAAATCTCAGATGAGCAACTGAGGAAAGATCTCT 1041  
251 GlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspGluLeu 267  
1042 GGTACCGAGAGAAATCTCTCAAAATGGAACAAAGTTGATGAACCTAG 1091  
267 ArgLeuGlnGlnGlnIleAspIleLysAlaThrGluArgSerThrLeu 284  
1092 AAGACTTCAGGACAAATGACATCAAGCAACAGAGAAAGTACATTAC 1141  
284 euThrGlnGlnHisGluLysLeuAlaIleAspSerGluGluAsnGluAsp 300  
1142 TTACGACACAGATGAAAGCTGCTGCACTTCTGAGGAAATTAAGAT 1191  
301 ThrAspGluGluLeuMetGluTrpIleThrLysPheGluGluArgIleAla 317  
1192 ACCGATGAGAACTAATGAAATGCAAAACAAATTTGAAGAAAGATGCT 1241  
317 AluLeuGluThrLysIleSerLysLeuValArgAspMetAspAspGlu 334  
1242 GTTACTAGAAACAAATCACTAACTGTGAAGATATGATGTGTAAG 1291  
334 lAspTyrSerSerValLeuSerLysGlnAsnSerGluLeuThrIleGlu 350  
1292 CATCTTAATAGCTCGTCTGTCCAAACAAATTTGAAATTAACATGAA 1341  
351 lIleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHisGlu 367  
1342 ATTGAAAGCTCCAGAGAGAGAGCTGATGCTACCTGATTAAGAAAGATGA 1391  
367 uArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeuGlyPro 384  
1392 ACGAGACTCAACATTAATAATATATGACATTAACATTAATCTTGCGCCG 1441  
384 AlProGluHisProPheThrAsnAspValAlaMetAsnLeuThrAsnArg 400  
1442 TTCCTGAACATCCCTTAACAAATGATGTTGCTATGAACTTCAACACAG 1491  
401 lLysAlaArgLeuSerSerLeuGluAsnAspLeuAspLysLysIle 417  
1492 ATTAAGGAGACATCAAGCTTGAGATGATTTGCTGGAATGAAGAA 1541  
417 sSerAsnGluAspGlnLeuAspValLeuTrpLysHisTyrLeuLysIle 434  
1542 ATCCAAATCAAGATCAGTTAGATGTTTGTGGAACACATCTTAATAATA 1591  
434 snAlaArgTyrSerGluValAspGlyGlnIleGlnSerLysIleGluSer 450  
1592 ATGCTCGCTACTCCGAAGTTGATGTCAGATACAACTAAGATTGAATCC 1641  
451 MetSerGlyIleLeuArgArgArgLysAspLysGluLysGluArgAspAl 467  
1642 ATGTCAGGCATTTTAAGCGGAGAAAGATTAAGGAAAGAAACGGAGTC 1691  
467 AlAlaGluValGluLeuSerLysPheAsnLeuSerArgIleAspGluArg 484  
1692 TGCAGAGTGAAGCTTTCAAAATTAATCTATCCGTTATCCATGAGAGG 1741  
484 lUArgHisMetGlnIleGluValGluArgLysThrLeuAlaLeuGlyGlu 500  
1742 AGAGACATATCAAAATTAAGTCAAGAGAGAGACACTGCGCTTGAGAA 1791  
501 ArgAspTyrAspSerIleIleSerGlnLysArgThrGluValTyrSerLe 517  
1792 AGAGCTATGATTCAAATTAATTAAGTCAGAACGACAGAGATTAATGTT 1841  
517 uGluGlnLysIleLysValLeuLeuArgGluLysAspIleIleAsnArg 534  
1842 GGAACAGAAATTAAGTCTCTGCGGAGAGAAATTAATTAATGA 1891  
534 snAlaAspGluArgValLysLeuGlyLeuLysLysAspAlaLeuGluSer 550  
1892 ATGCTGATGAAGAGATTAAGTCTGCTGGAAGAGATGATGGAAGC 1941  
551 SerLysAspLysLeuAsnGlnIleValAsnGlnHisLysAspLysIleLys 567  
1942 AGCAAGGACCAAGCTCAATGAGATGATTAAGTCAAGTAAATGA 1991  
567 sLysValLeuArgGluArgAsnProPheGluLysAspMetLysLysGlu 584  
1992 AAAGTACTTGGGGGAGGATCTCTTGGACAGAGATTAAGAAAGGAGA 2041  
584 lAsnGlnAlaPheTrpProValAspLysGluTyrAsnGluLeuArgSer 600  
2042 TCAATCAAGCTTTGGCTGGGACAGAGAAATCAATGAGATTGAATCA 2091  
601 LysSerGlnGluAlaGluGlnGluLeuLysPheThrGlnSerLysValThr 617  
2092 AAATCCAGAGAGAGAGAGAGAGCTTAATTAATTTACTCAGAGCAAGTAC 2141  
617 rAspAlaArgGluGlnLeuThrLysLeuArgArgAspMetAspAlaLys 634  
2142 TGATGCTGAGACATTAATGACAAATCTCGAAGAGATATGATGCAAAA 2191  
634 rArgPheLeuAspSerLysLeuGlnSerIleLeuGlnIleSerAlaAsn 650  
2192 GAGATCTCTGAGCTGAAACTCAATCTAATTTTACAGATATCTCTAAT 2241

651 ValAspMetPheProLysValLeuGlnAspAlaMetAsnLysArgAspG1 667  
|||||  
2242 GTTGACATGTTTCCCAAGTTCTACAAGACGCATGACAAAGAGATGA 2291  
667 uGlnLysArgLeuGlnAsnPheAlaAsnGlyMetArgGluMetLeuAlaP 684  
|||||  
2292 ACAGAAAAGATTAGAGAAATTTCCAAATGGAAATGCGGAAATGCTTGCA 2341  
684 roPheGlnHisLeuAlaArgLysAsnHisValCysProCysGlyArg 700  
|||||  
2342 CTTTGAACATTTGGCTCGAAGAAATCATGTAGCCCATGCTGTGAACGT 2391  
701 AlaPheThrProAspGluGluAspGluPheValLysGlnArgMetG1 717  
|||||  
2392 GCTTTACACCTGATGAGAGAGATGAGTCTGTGAAGAAAGATGCA 2441  
717 nAsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSer 734  
|||||  
2442 AAACCTCAAGTACTGCAGAGAGATCTAAAGCTCTGCAATGAAATCATCA 2491  
734 snAlaGluAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyrAsp 750  
|||||  
2492 ATGCTGAACCTCTTTTTCAGCAATGGATTAACCTTGACATCTATGTAT 2541  
751 AlaTyrValLysLeuValGluGluThrIleProLeuAlaGluLysAsnLe 767  
|||||  
2542 GCTTATGTAAGCTGTTAGAAAGAACCATCTCTAGCAGAGAAAACCTT 2591  
767 uAsnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaPheAspAspL 784  
|||||  
2592 GAATCAACATTTGGCGATGAAAGTCAAGAGCGCAGCATTTGATGATC 2641  
784 euLeuGlyValLeuAlaHisValGlnMetAspArgAspAlaValGluAla 800  
|||||  
2642 TTTTGGGTCTTCCCATGTTCAATGACAGAGGATGACAGTGAAGGCC 2691  
801 LeuLeuGlnProThrAspThrIleAspArgHisValHisGluIleGlnG1 817  
|||||  
2692 TTATTAACAACCCATGATCTATGACAGCATGTACATGAATTTACA 2741  
817 nLeuValLysGluValGluAspLeuGlnLysIleValAlaLeuAspSerSerGly 834  
|||||  
2742 GCTAGCAAGAAGTAGAGATCTTGAATATGCACTTGATTTGATGAGCC 2791  
834 rGlyValLysSerLeuGlnGluIleGlnLeuGlnLeuAsnPheLeuGln 850  
|||||  
2792 GAGGTGTCAAGCTTTGGAGAAATTCACACTGAGACTGAACCTTTCGCA 2841  
851 ArgThrArgAspThrLeuIleValGluValAspAspLeuArgAspGlnH 867  
|||||  
2842 AGAACCAAGGCACATGTATGTGCAAGTGATGATCTTAGAGATCAACA 2891  
867 sArgMetLeuAsnGluAspMetSerSerAlaGlnValArgTyrHisAsnA 884  
|||||  
2892 TAGAATGCTGAATGAAGATATGCAAGTCTCAGTGAGATGGCAAAAG 2941  
884 lArgGluLysValLysAlaSerSerIleLeuGlnArgPheGlnLys 900  
|||||  
2942 CTCGGGAGAGAAAGTGAACCTTCTAGCATTTGGAAAGATTCCAAAA 2991  
901 SerGluGluGluLeuValLeuLeuAlaGluGluLysGluGlnLeuLeVal 917  
|||||  
2992 TCTGAAGAGAAATGTGCTTCTACCTGAGAAAAGAACCAACTGATTT 3041  
917 lGluLysLysLeuLeuGluGluSerLeuAspProLeuSerLysGluLysG 934  
|||||  
3042 AGAAAAGAGCTTTAGAAAGATCTCTTATCATTTGTCCAAAGAGAAAG 3091  
934 lSerLeuLeuGlnLysTyrAsnAlaLeuLysGlnLysLeuAspGluGlu 950  
|||||  
3092 AGAGCTTGTTCAGAGATATATGCTTGAACCAAAAGCTGATGAAGAG 3141  
951 TyrHisGlnLeuAlaGluArgLysArgGluPheGlnGlnLeuAspAl 967

|||||  
3142 TATCATACGTTTCAGAAAAGAAAAGGAGCTTCACACGAACCTTATG 3191  
967 AluGlyArgLeuAsnMetLysIleLysGlyTyrLeuAspSerLysAla 984  
|||||  
3192 TCTTGGAAAGACTTAATATGAAGATTAAGAGGTACTTGATTCACAAAA 3241  
984 snGluLysLeuLysGluLeuGlnGlyArgHisValLeuCysHisSerGln 1000  
|||||  
3242 ACAGAAAGCTTAAGGAATTCAGAGGAGGAGCATTTCTTCCATTCAG 3291  
1001 LeuGlnSerCysMetAlaLysGlnGlnArgIleSerAlaGluLeuAsnLys 1017  
|||||  
3292 TTACAGAGTTGCTGCGAAGAACAGCAAGAAATATCAGCTGAGTTAAC 3341  
1017 sSerLysGluLeuLeuGlnGlyGlnGlnLysGlnLysArgAsnIleAspA 1034  
|||||  
3342 GAGCAAGAAGACTACTGACAGGCGCCAGGCGCAGTTGAAGAAACATGAG 3391  
1034 sPsnLeuLysTyrArgLysThrLysAlaAspValGluGlnLeuThrArg 1050  
|||||  
3392 ACAATCTCAAGTACAGAGAAACAAAGCTGATGTGGAACACTTACTCGT 3441  
1051 AspIleGluSerLeuGlnGluArgLeuLeuSerIleGlySerLeuSerAl 1067  
|||||  
3442 GATATGAATCACTTGAAGAAAGCTGCTTTCATATGATGCTGTCTGC 3491  
1067 AlieGluAlaAspLeuLysArgHisSerGlnGlnLysGluArgLeuAsn 1084  
|||||  
3492 TATAGAGCGATCTGAAGGCGCATCTTCAGAAAGAAAGAGAGCTTAAT 3541  
1084 eArgLysAsnArgTyrGlnGlyThrLeuSerValTyrGlnSerAsnIle 1100  
|||||  
3542 CAGAAATTAACAGGTGCAAGGAACACTTCTGTTATCAAGATATAT 3591  
1101 SerLysHisLysGlnGluLeuLysLysSerGlnTyrLysAspIleGluLys 1117  
|||||  
3592 TCAGACACCAACAGAGAGCTTAACTGTCACAGTACAGAAATTCAGAA 3641  
1117 sArgTyrThrAsnGlnPheLeuGlnLeuLysThrThrGluMetAlaAsnL 1134  
|||||  
3642 GCGATATCAATCAATTTCTCAGCTTAAGCAACTGAAAGGCAAAACA 3691  
1134 yAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPhe 1150  
|||||  
3692 AGGACTTGACAGATATATATGCTTATAGCAAGGCTCTTATGCGGCTC 3741  
1151 HisSerMetLysMetGluGluLysIleLysLysIleIleGluLeuThrG1 1167  
|||||  
3742 CACAGCATGAAGATGAGGAGATTAATTAATCAACGAACGTGCGCA 3791  
1167 nGlnThrTyrArgGlyGlnAspIleAspTyrIleSerIleAsnSerAspS 1184  
|||||  
3792 ACGAGATACAGAGCGCAGATATGATTAACATMAACATTAATTTGAT 3841  
1184 eArgLysAlaGlyThrArgSerTyrSerTyrArgValAlaMetGlnThr 1200  
|||||  
3842 CTGAGGCTGCTGCACTCGATCATACGCTACCGGCTTGTATGCAACT 3891  
1201 GlyAspAlaGluLeuGluMetArgGlyArgCysSerAlaGlyGlnLysVal 1217  
|||||  
3892 GGTGATGCTGAGCTGGAATGCGAGGCGCTCAGGCTGCGTCAAGAGT 3941  
1217 lLeuAlaSerLeuIleIleArgLeuAlaLeuAlaGluThrPheCysLeuA 1234  
|||||  
3942 TCTGTGTTCTTATATACAGCTACGACTTGCAGAACTTCTGCTGCA 3991  
1234 snCysGlyIleLeuAlaLeuAspGluProThrThrAsnLeuAspGlyPro 1250  
|||||  
3992 ACTGCGGTATATGCTTGTGATGAGCCAACTACCAATCTAGATGGCCA 4041  
1251 AsnAlaGluSerLeuAlaAlaLeuLeuArgIleMetGluAlaArgLys 1267  
|||||

```

4042 AATGACAGAGTCTGCTGCGCTGTTGAGAAATATGAGACGACGAA 4091
1267 gcygnglnuansphnglnleuilefalllethrshaspguarugphea 1284
|||||
4092 AGGCGACAGAAACCTCCAGTTGATTTAAVCACTATGATGAGCAATTGG 4141
1284 lahlseuilegylgnatrglnleualagllysteytrayrvalser 1300.
|||||
4142 CCCATCTTATCGGTCAAGGACGCTTGCAGAAAGATGATGAGATCGCC 4191
4192 AAGGATGAGAACCGACGACACCATATATGATCCCAAGAGATTTTGGAC 4239
seq_name: /STDJL/gcgdata/geneseq/geneseq/NM1997.DAT:AA75237
seq_documentation_block:
ID AA75237 standard; CDNA to mRNA; 5893 BP.
XX
AC AA75237;
XX
DT 21-DEC-1998 (first entry)
XX
DE Nucleotide sequence encoding human RAD50.
XX
db: human; RAD50; DNA repair; tumour suppression; cancer; Septin-2;
KM central nervous system.
XX
XX Homo sapiens.
XX
Key Location/Qualifiers
FT CDS 389..4327
FT /tag= a
FT /product= "RAD50"
XX
XX MO9727284-A2.
XX
XX 31-JUL-1997.
XX
XX 24-JAN-1997; 97WO-US01299.
XX
XX 17-JUL-1996; 96US-0687080.
XX
XX 26-JAN-1996; 96US-0592126.
XX
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX Dolganov G;
XX
XX WPI; 1997-393672/36.
XX
XX P-PSDB; AAW2275.
XX
XX Human tumour suppressor gene RAD50 - useful to detect
XX predisposition to, decrease risk of and treat cancer; also Septin-2
XX homologues
XX
XX Claim 4; Page 73-75; 195pp; English.
XX
XX The human RAD50 (hRAD50) is involved in DNA repair and has tumour
XX suppression activity, can be used to detect predisposition to, decrease
XX the risk of or treat cancers, e.g. acute myeloid leukaemia,
XX myelodysplastic syndrome, therapy related myelodysplastic syndrome,
XX anemia related acute myeloid leukaemia, refractory anaemia or refractory
XX Septin-2 homologues of which may be used as targets for cancer therapies
XX and central nervous system directed treatment methods, and to measure the
XX proliferative potential of selected cell types.
XX
XX Sequence 5893 BP; 2081 A; 1097 C; 1301 G; 1414 T; 0 other;

```

```
alignment_scores:
  Quality: 1730.00
  Ratio: 1.918
  Length: 1385
  Gaps: 38
```

Percent Similarity: 65.126      Percent Identity: 32.415

alignment\_block:  
ITS-09-538-396-2 x AAT75237

05-09-538-396-2 X AAT/5237

Align seg 1/1 to: AAT75237 from: 1 to: 5893

1 MetSerThrValAspLysIleuLeuIleuLysGlyIleIleArgSerPhe..... 15  
389 ATGTCCCGGATGCAAAAGATGACATCTTCGGCGTGGCAGATTGGAAAT 438  
16 AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuT 32  
439 AGAGGCAAAAGATTAACCAAAATTATCTCTTCTTCACGCCCTTACAAATT 488  
32 IeValGlyProAsnGlyAlaGlyLysThrThrIleIleGlyCysLeuLys 48  
489 TGGTTGACCCCAATGGGGCGGGAAACAGCACATCATGATGCTMAAA 538  
49 LeuSerCysThrGlyCysLeuLeuProProAsnSerArgSerGlyIleThrP 65  
539 TATATTATTGACAGCAATTTCCCTCCTGGAAACCAA...GGAAATACATT 585  
65 eValHisAspProLysValAlaGlyLeuThrGluThrLysGlyGlnIleL 82  
586 TGTACACGATCCCAAGCGTGTCTCAAAACAAAGATGTGAGAGCCAGATTC 635  
82 YsLeuArgPheLysThrIleAlaGlyLysAspValValCysIleArgSer 98  
636 GCTCGCAATTTCTGTATGTCATGACGAGCAACTTATTAGCTGTGCAAAAGATC 685  
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSe 115  
686 ATGGTGTGACTCAGAAAGCAAAAGACAGAAATTMAACTGTGGAGAG 735  
115 rValLeuGlnThrIleAsnProHisThrGlyGlyLysValCysLeuSerT 132  
736 AGTCATT...ACTAGAACAAAGCAT...GGTGAAGAGTCAGTGTAGCT 779  
132 yArgCysAlaAspMetAspArgGluIleProAlaLeuMetGlyValSer 148  
780 CTAAAGTGTGAGAAATTGACCGAGAAATGATCATGCTCTTGGGGGTTTC 829  
149 LysAlaValLeuGlnLysnValIlePheValHisGlnAspGluSerAsnTr 165  
830 AAGGCTGTGTAATATGTCATTCTTGTCTCATCAACAAAGATTGTAATG 879  
165 pProLeuGlnAspProSerThrLeuLysLysLysPheAspAspIlePheS 182  
880 GCCTTTAAGGAGGAAGAGCGCTTGAAGCAAAAGTTGAGAGATTTTTT 929  
182 eValThrArgTrpThrLysAlaLeuGlnValIleLysLysLeuHisLys 198  
930 CAGCAACAGAGATCATTAAGCCCTTAAACACACTTCGCGAGGTGCTGAC 979  
199 AspGlnMetGlnGluIleLysTrpPheArgLeuLysLeuGlnAsnLeuG 215  
980 ACACAAAGGTGAGAAAGTAAAGATATCAAAATGCAACATTAATCTGAA 1029  
215 nThrValLysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAspG 232  
1030 GCAATTAATAGGAAAAAGCTGTGAGATCTGTGATCAAGATTAACAAGTAAAG 1079  
232 InoLysSerAspAlaSerLysSerGlnMetGlnLeuGlnLeuLysGluLys 248  
1080 AACCACCAAGTTAAACATCTTCAAAGAGAAATTGTCAAATCTTATGAAATGAA 1129  
249 IleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspG 265  
1130 CTTGATTCATTAAGAAATCGCTTAAAGCAAAATTAACATATATCTCTGAA 1179  
265 uLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSerT 282

```
1180 AATAATGAACCTTGACATGAAATTAAGCCTTGATAGCCGAAGAGC 1229
282 htleuethrGInGInHISGLuLysLeuAlaLeuSerGluGluAsn 298
1230 AATGGAGAAAGATTAATAGTGAAGCTGGAAGAAAGAAAGGTTT 1279
299 GLAsPThrAspGluGluMetGluTrpGluThrLysPheGluGluAsp 315
1280 CAAGGAGCTGATGACAACTAAATGACTTATATACAAATCAGAGAAC 1329
315 gtleAlaLeuLeuGluThrLysLysSerLysLeuValArgAspMetAsp 332
1330 AGTAAGGAGAAAGAAAGAAATGGTAGAGCTGATCGTGAAGCTGAAA 1379
332 spGluAlaSerTyrSerSerValLeuSerLysGlnAsnSerGluLeuThr 348
1380 AACTAAATAAAGAAATAGCTTCTCAATCAGAAATCAGAAATCTGCTT 1429
349 HlSGluLleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThrMetLys 365
1430 GTTGAACAGGCTGCTACAGCTGCAAGAGATCCCATCAGACACTAT 1479
365 sHISGluArgAspSerAspLleLysAsnLecTyrThrLysHisAsnLeuG 382
1480 CCGAGCTAGAGATTCATTATTAATCAGCTTTGGCAACACAGCTAGAA 1529
382 LyrProValProGluHisProPheThrAsnAspValAlaMetAsnLeuThr 398
1530 ATGGCTTTGAGCGTGCATCTCAGTGAAGACAGATTAATAATTTTCC 1579
399 AsnArgLleLysAlaArg.....LeuSerSerLec 408
1580 AAACCTGTGAGAGAGAGACAGAGGAGCAAAAACTGCCAACCAACT 1629
408 uGluAsnAspLysLeuAspLysLysSerAsnGluAspGlnLeuAspV 425
1630 GATGAATGACTTTCAGAAAAAGAGACTGGAACAAACAAACAGATGAT 1678
425 alLeuTrpLysHisTyrLeuLysLleAsnAlaArgTyrSerGluValAsp 441
1679 .....GAGTAAAGATTAAGAAAACTGAGCTGGA 1708
442 GlyGlnLleGlnSerLysLleGluSerMetSerGlyLleLeuArgArg 458
1709 AGAATTAATGAGTTAAATCAGAAATCTTAAGT.....AA 1743
458 gLysAspLysGluLysGluArgAspAlaAlaGluValGluLeuSerLys 475
1744 GAAGCAGATGAGCTGAAAAATGTGAAGTATGAATACACAGCTGGAAG 1793
475 heAsnLeuSerArgLleAspGluArgGluArgHisMetGlnLleGluVal 491
1794 GATCTTCAGACAGATTTCTGAACTGAGCAGAGAGCTC...ATAAAGCT 1840
492 GluArgLysThrLeuAlaLeuGlyGluArgAspTyrAspSerLleLec 508
1841 GAACGTGAG.....TTAAGCAAGCGTGAAGAAAAACAGCAATGTGA 1881
508 rGlnLysArgThrGluValTyrSerLeuGlnGlnLys..... 520
1882 AACCTTAATAAATGAAGTAAATAGTCTCCAAATGAAAAAGCAGACTAG 1931
521 .....LleLysValLeuLeuArgGluLysAspLleLleAsnArgAsn 534
1932 ACAGAGACCTCGCTAAACTGACAGAGATGAGCAGCATTAAACCAT 1981
535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGlu 550
1982 ACACAACACGTACCCAAATGAGATGCTGACCAAGACAAAGCTGCAA 2031
550 rSerLysAspLysLeuAsnGluLleValAlaGlnHisLysLysLleL 567
2032 AGAT...GAACAATCAGAAAAATTAATCTAGCACAGTGAATTA 2078
567 yLysValLeuArgGlyArgAsnProPheGluLysAspMetLysGlu 583
2079 CTTCACTGTTG...GGATATTTTCCACAAAAAGCGCTGAAGAC... 2122
584 lleAsnGlnAlaPheTrpProValAspLysGluTyrAsnGlnLeuArg 600
2123 .....TGG.....CTACATAG 2133
600 rLysSerGlnGluAlaGluGlnGluLeuLysPheThrGlnSerLysVal 617
2134 TAAATCAAAAGAAATATATCAG..... 2155
617 htrAspAlaArgGluGlnLeuThrLysLeuArgArgAspMet..... 630
2156 .....ACCAGGACAGCTTGGCCAAATGAAACAGAGAACTAGCTTCA 2200
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerLleLeuG 646
2201 GACGAGAAATAAATCATATTAATTAATGAACATAAAGAAAGAGAGCA 2250
646 nleSerAlaAsnValAspMetPheProLysValLeu.....GlnAsp 661
2251 GTTGTCAGATTACGAAAGACAGCTGTTGATGTTGTGTGAGCAGAGAT 2300
661 lAmetAsnLysArgAspGluGlnLysArgLeuGluAsnPheAlaAsnGly 677
2301 TTGAAGTGATTTAGACAGCTTAAAGAGAAATTAATAATCATCAAAA 2350
678 MetArgGluMetLeuAlaPro.....PheG 686
2351 CAGGAGCCATGCTGCTGAGGACACAGAGCTTACTCCAGTTCATTAC 2400
686 uHisLeuAlaArgLysAsnHis...ValCysProCysGluArgAlaP 702
2401 TCAGCTACAGAGAAACAGCATGTTGCCCGTTTCAGAGAGTTT 2450
702 he...ThrProAspGluGluAspGluPheValLysLysGlnArgMetGln 717
2451 TTCAGACAGAGCGCTGATTAAGAAAGTCACTGATGATTTGCGTCTAA 2500
718 AsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSer 724
2501 CTGCGACTTCTCCAGTAACTCAAGTCAACAGAAATCAGAGCTAAAA 2550
724 nAlaGluAlaLeuPheGlnGlnLeuAspLysLysLeuArgThrLleTyrAsp 751
2551 AAAGAAAGACGGCGTGAATAATGCTGGAGCTTGCCCATGAGGCAA 2600
751 lArGlyValLysLeuValGluGluThrLlePro...LeuAlaGluLys... 765
2601 GCATTAATTTGATTTGAGAGAGAAATACCAAAATTAAGAAACAACTG 2650
766 ...AsnLeuAsnGlnHisLysAlaAspGluSerGlnLysAlaGlnAlaP 781
2651 CAGATGTCATATAGACATACAGCGCTTAAGAGACATAGAAAGACA 2700
781 eAspAspLeuLeuGly..... 786
2701 AGAAACACTTGGGTACATTAATGCTGAAGAAAGAAAGTCCAAAGTAT 2750
787 .....ValLeuAlaHisValGlnMetAspArgAspAla 797
2751 GCCTGACAGATGTTACAAATTAATGAGAGGTTCCAGATGCAACTAAAGAT 2800
798 ValGlu.....AlaLeuGlnProThrAsp 807
2801 GTTGAAGAAATAATGACACACAGAGCTTAACCTACAGAGAAATAGC... 2848
807 rleAspArgHisValHisGluLleGlnGlnLeuValLysGluValGlu 824
2849 .TAGATGCAAGCTGCCAACAGTCAACAG...GAAAAACAAG 2888
```

```

824 sprleuGlutylAlaLeuaspSerSerGlyArgGlyValIysSerLeu 840
      ::::::::::::::::::::
2889 AGAAGACGACCAAGTATGACACA.....GTTCTAGTAATAT 2926
      |||
841 GlutIleuGluLeuAsnPhleuGlnArgThrArgAspThrIleu 857
      |||
2927 GAATTGAATGCTAGCTT..... 2944
      |||
857 eValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGluAsp 874
      ::::::::::::::::::::
2945 .....ATACGAGCACGACGAGAACAGATTCACATCTAA 2978
      |||
874 etSerSerAlaGlnValArgThrHisAsnAlaArgGluGluValIys 890
      |||
2979 AAAGTACAAACA.....AATGAGCTAAATCTGAGAACTTCAG 3016
      |||
891 AlaserSerIleuGluArgPheGlnIlySerGluGluValIle 907
      |||
3017 ATATCCACTAATTGCAACGTCGACGAACTGAGGACGACTGTGGA 3066
      |||
907 uLeuAlaGluGluIlySerGluGlnIleValGluIlySerLeuGlu 924
      |||
3067 ATATCCACTGAGTCTGCTTGTACAGAGATTAAGATGCTAAG 3116
      |||
924 luserLeuAspProLeu.....SerIlyGluIys 933
      |||
3117 AGCAGGTAGGCCCTTGGAAACCAATTGGAAAGTCCACCAAGAAA 3166
      |||
934 GluserLeuGluGlnIlyArgAsnAlaLeuIlySerGluIlySerAspGlu 950
      |||
3167 GAGATTAATCAACAACAAATAATACAGCAACAAATAGCACAGATTA 3216
      |||
950 uTrpHisGlnLeuAlaGluArgIlyArgGluPheGlnGlnIleuAsp 967
      |||
3217 ACTGATATGATTAAGAGGTTAAATAATTCATGCTATATG... 3262
      |||
967 lAlaGluArgLeuAsnMetIlySerGlyTrpLeuAspSerIly 983
      |||
3263 .....AAAGCATGTGAATTTATTCAGAGTGGAAA 3295
      |||
984 AsnGlu.....LysLeuIlySerGluLeuGlnIlyArgHisValIleu 998
      |||
3296 GACGACTAATAGAACCAAAAGAACTGACTTAATAAGTAAAT... 3340
      |||
998 sSerGlnLeuGlnSerCysMetAlaIlyGlnGlnArgIlySerAlaGlu 1015
      |||
3341 .GCTCACTAAGTGAATGCGAGAAACACAAAGAAATTAATGAAGTA 3389
      |||
1015 euAsnIlySerIlySerGluLeuGlnGlnIlyGlnIleuIlyArgAsn 1031
      |||
3390 TGAGACTCATGACAGACAGATNTGTACACAGACAGATACAGAAAGGTGG 3439
      |||
1032 lIleAspAsnIleuIlySerGlyArgThrIlyAsnValIleGluGln 1048
      |||
3440 CTACAGATTAATCTTAAAGAAAGAAATGAGAACTTAAGAGAGT 3489
      |||
1048 uThrArgAspIleuSerLeuGluArgGluLeuSerIleGlySerL 1065
      |||
3490 TGA.....GAGAAAGAAACCAACATTTGAAGGAAATGGGTCAA 3530
      |||
1065 euSerAlaIleGluAlaSerLeuIlyArgHisSerGlnGluIys... 1079
      |||
3531 TCCAGCTTTGCAATGAAAGTGAACATCAGAACTTGAGAGAGACATA 3580
      |||
1080 GluArgLeuAsnSerGluPheAsnArgTrpGlnIlyThrIleSerValIy 1096
      |||
3581 GACATATATAAAGAAATCATATTTGGCATTAAGGCGACAGAAAGGTAA 3630
      |||
1096 rGlnSerAsnIleSerIlyHisIlyGlnGlnIleuIlySerGlnIly 1113
      |||
3631 TGAAGAGAAATTAATTCATTTTAAGAAAGAACTTGAGAACCAATTC 3680
      |||
1113 yAspIleuIlyArgIlyThrAsnGlnPheLeuGlnIleuIlySerThr 1129

```

```

      ::::::::::::::::::::
3681 GCGATGCTGAGAGAAAGTATAGAGAAATGATGATTGTATGAGGACACA 3730
      |||
1130 GluMetAlaAsnIlyAspLeuAspArgTrpTrpThrAlaLeuAspIly 1146
      |||
3731 GAACCTTGAAACAAGAGATGATGATATTTATTAAGACTCTGACCAAGC 3780
      |||
1146 AluMetArgPheHisSerMetIlySerMetGluGluIleAsnIlyIle 1163
      |||
3781 AATATGTAATTTCACTATGAAATGAAATGAAATCAATTAATTTATC 3830
      |||
1163 ysgIleuTrpGlnGlnThrTrpArgGlyGlnAspIleAspTrpIleSer 1179
      |||
3831 GTGACCTGTGGCGAAGTACCTATCTGACCAAGATATTGATACATAGAA 3880
      |||
1180 lLeuAsnSerAsp.....SerGluIlyAlaGlyThrArgse 1191
      |||
3881 ATACGGTCTGATGCCGATGAATAATGTATACGCTCTGTAAAGCCGAA 3930
      |||
1191 rTrpSerTrpArgValIleMetGlnThrGlyAspAlaGluLeuGluMet 1208
      |||
3931 TTATTAACCTACGAGTGTGATGTGAGAGGAGACACACCTTGATATGC 3980
      |||
1208 rGlyArgCysSerAlaGlyIlyIlyValIleuAlaSerLeuIleIleArg 1224
      |||
3981 GAGGACGATGACGTGCTGACAAAGATATTACCTCATCATCTGC 4030
      |||
1225 LeuAlaLeuAlaGluThrPheCysLeuAsnGlyIlyIleuAlaLeuAs 1241
      |||
4031 CTGGCCCTGCTGAACCTTCTGCTCAACTGTGACATCATTCCTGGA 4080
      |||
1241 pGluProThrThrAsnLeuAspGlyProAsnAlaGluSerLeuAlaAla 1258
      |||
4081 TGAGCCAAACAATCTTGTGACCGAGAAACATGAACTCTTGACACTG 4130
      |||
1258 lAlaLeuArgIleMetGluAlaArgIlyGlyGlnGlnIleuAsnPhleu 1274
      |||
4131 CTCGGTGAATATAAAGTCCGACAGACAGCCTCACTCAGCTT 4180
      |||
1275 lIleValIleThrHisAspGluArgPheAlaHisIleuIleGlyGlnArg 1291
      |||
4181 CTGTAATCACTCATGATGAAGATTTGTGGCTTTAGACGCTTGA 4230
      |||
1291 nLeuAlaGluIlyTrpTrpArgValSerIlyAspGluAsnGlnHisSer 1308
      |||
4231 ATATGTGAGAAATTTCTACAGATTAATAAGAACTCATCATCTGCTCAG 4280
      |||
1308 lIle 1309
      |||
4281 AGATT 4285

```

seq\_name: /SDSL/gcgdate/geneseq/geneseqn/NA1998.DAT:AAV59979

seq\_documentation\_block:

ID AAV59979 standard; cDNA to mRNA; 5893 BP.

XX

AC AAV59979;

XX

DT 25-NOV-1998 (first entry)

XX

DE Nucleic acid encoding a human homologue of yeast RAD50.

XX

KW Human homologue; yeast RAD50; Drosophila Septin-2; Acyl-CoA synthetase;

KW Immunomodulatory activity; identification; activated T-cell; ds.

XX

OS Homo sapiens.

XX

PH Key

FT CDS 389..4325

XX /tag= a

PN W09838306-A1.

XX



425 alleuTPRLysHISlytleuLysIleasnalaAgyTYSerGIuValasp 441  
1679 .....GAGATAGAGATAGAAAATGAGCTGAGGGA 1708  
442 GlyInIleGInserLysIleGInserMetSerGlyIleuArgLysArg 458  
1709 AGATATATGAGTTAAATCAGAAAATCCTTAAGT .....AA 1743  
458 GlyAspLysGIuLysGIuArgAspAlaIleValGIuLeuSerLysP 475  
1744 GAGAGAGATAGAGCTGAGAAAATGAGATATAGATTAACAGCAGTTGAGAG 1793  
475 heAsnLeuSerArgIleAspGIuArgGIuArgHISMetGInIleGIuVal 491  
1794 GATCTTCAGACGAGATCTTGAATGAGCAGCAGGAGCTC...ATAAAGCT 1840  
492 GIuArgLysThrLeuAlaLeuGIuLysArgAspLysSerIleIleSe 508  
1841 GAACTGTAG.....TTAAGCAAGCTGAGAAAACAGCAATGTAGA 1881  
508 rGIuLysArgThrGIuValIYrSerLeuGIuLys..... 520  
1882 AACCTTAATAAGAGAAATAGATCTCCAAAATGAAAAGCAGACTTAG 1931  
521 .....IleLysValIleuLeuArgGIuLysAspIleIleAsnArgAsn 534  
1932 ACAGGACCTGCTGAACCTGACAGGAGATGAGCAGCTTAACATCAT 1981  
535 AlaAspGIuArgValLysLeuGIuLysLysAspAlaLeuGIuSe 550  
1982 ACACACACGCTACCCAAATGAGATGCTGACCAAGACAAAAGCTGACAA 2031  
550 rSerLysAspLysLeuAsnGIuIleValAsnGIuHISLysAspLysIle 567  
2032 AGAT...GACAAATCAGAAAATTAATCTAGCAGCAGTGTAGATTA 2078  
567 yALysValIleuArgGIuLysArgAsnProPheGIuLysAspMetLysLysGIu 583  
2079 CCTCACTGTG...GGATTTTCCCAACAAAACAGCTTGAAAGC... 2122  
584 IleAsnGIuAlaPheTrpValAspLysGIuLysArgGIuLeuArgSe 600  
2123 .....TGG.....CTCAATAG 2133  
600 rLysSerGIuIleValGIuLysGIuLeuLysPheThrGInserLysValT 617  
2134 TAAATCAAAAGAAATTAATCAG..... 2155  
617 hAspAlaArgGIuLysIleuThrLysLeuArgArgAspMet..... 630  
2156 ....ACGAGGACAGACTTGCCTCAATTAAGCAAGAACTAGCTTCATCT 2200  
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GInserIleLeuGI 646  
2201 GAGCAGAAATAAATCATATAATAGAACTAAAGGAGGAGGAGCA 2250  
646 nIleSerAlaAsnValAspMetPheProLysValLeu.....GInAspA 661  
2251 GTTGTCACTTACAGACAGCAGCTGTTGATGTTGTGTGACCAAGATT 2300  
661 IAsnLeuLysArgArgGIuLysArgLeuGIuLysArgPheAlaAsnGI 677  
2301 TTGAATGATTTAGACAGCTTAAGAGAAATGAAAATCAATCAAAA 2350  
678 MetArgGIuMetLeuAlaPro.....PheGI 686  
2351 CAGCGAGCAGCTGCTGCTGAGCCACAGCAGTTTACTCCAGTTCAATAC 2400  
686 uHISLeuAlaArgLysAsnHIS...ValLysProCysCysGIuArgAlaP 702  
2401 TCAGCTTAACAGAGAAAACAGTCATGTTGCCCGTTTGTCAAGAGATT 2450

702 he...ThrProAspGIuLysArgPheValLysLysGIuArgMetGIn 717  
2451 TTCAGACAGAGCTGAGTTACAGAAAGTCAGTCACTTTCAGCTTAA 2500  
718 AsnSerSerThrAlaGIuArgSerLysAlaLeuAlaMetGInSerSerAs 734  
2501 CTGCGACTGTCTCCAGATTAACCTCAGTCACAGCAAGATCAGAGCTAA 2550  
734 nAlaGIuAlaLeuPheGInGInLeuAspLysLeuArgThrIleLysAsp 751  
2551 AAAGAAAAGCGCGGTGTAATGCTGGAGCTTGCCCATGAGGCATA 2600  
751 IArgValLysLeuValGIuLysThrIlePro...LeuAlaGIuLys... 765  
2601 GCATTAATGATTTAGAGAGAGAGAAATCCAGATTAAGAAACAACTG 2650  
766 ..AsnLeuAsnGInHISLeuAlaAspGIuSerGInLysAlaGInAlaP 781  
2651 CAGATGTCAATAGACATACAGCGCCTAAAGAACATAGAGACAA 2700  
781 eaSPAspLeuLeuGIuLys..... 786  
2791 AGAAACACTGTGGTCAATATGCTGAGAGAAAGAAAGTCCCAAGATT 2750  
787 .....ValLeuAlaHISValGInMetAspArgAspAla 797  
2751 GCCTACAGAGTTCAATTTATGAGAGAGCTTCCAGATGGAATTAAGAT 2800  
798 ValGIu.....AlaLeuLeuGInProThrAspTh 807  
2801 GTTGAAGAAAATTCACAAACACAGCTTAAGCTAACAGAAATGAGAC... 2848  
807 rIleAspArgHISValHISGIuIleGInLeuValLysGIuValGIuAla 824  
2849 ..TTAGATGAACGTGTCCAAACAGTCAACAG.....GAGAAACAG 2888  
824 sPLeugIuLysAlaLeuAspSerSerGlyArgGIuValLysSerLeuGIu 840  
2889 AGAAACAGCAGATTTAGACACA.....GTTTCTGTAGTAAAGATT 2926  
841 GIuIleGInLeuGIuLeuAsnPheLeuGInArgThrArgAspThrLeuI 857  
2927 GATTTGAATCTGAACCT..... 2944  
857 eValGIuValAspAspLeuArgAspGInHISArgMetLeuAsnGIuAspM 874  
2945 .....ATACAGCAGCAGCAGACAGATTCACATCTAA 2978  
874 eTSerSerAlaGInValArgTPHISAsnAlaArgGIuLysValLys 890  
2979 AAAGTACACA.....ATGAGCTTAATTTGAGAACTTCAG 3016  
891 AlaSerSerIleLeuGIuArgPheGInLysSerGIuGIuLysValLe 907  
3017 ATATCCACTAATTTGCAACGTGCTCAGCAACTGAGAGACAGACTGTGA 3066  
907 uLeuAlaGIuGIuLysGIuLysLeuIleValGIuLysLysLeuLeuGIu 924  
3067 ATTATCCACTGAGTTCAGCTTTGTACAGAGATTAAGATGCTTAAG 3116  
924 IAsnLeuAspProLeu.....SerLysGIuLys 933  
3117 AGCAGGTAAAGCCCTTTGCAACACACTTGGAAAAGTTCACAGCAAGAAA 3166  
934 GIuSerLeuLeuGIuGIuLysArgAlaLeuLysGIuLysLeuAspGIuGI 950  
3167 GAAAGATTAATACACAAAATAATACAGCAACAAATAGCAGAGATTA 3216  
950 uTYrHISGInLeuAlaGIuArgLysArgGIuPheGInGInGIuLeuAspA 967  
3217 ACTGAATGATTAATAAGAGAGTTTAAATATTCATGCTTAATG... 3262  
967 ILeuGIuArgLeuAsnMetLysIleLysGIuTYrLeuAspSerLysLys 983

```

3263 .....AAAGACATGAGATATATATCAAGATGGGAAA 3295
984 AsnGlu.....LysLeuYsgLUleuGInGlyArHIsValLeuCySH 998
3296 GACGACTATAGAGCAAAAGAACTGAATTAATAGTATA..... 3340
998 sserGluLeuGInSerCyMeValAlYsgInGInArGILeserAlaGluL 1015
3341 .CTCAACTAGTGAATGCGAGAACACACAAAGAAAGATTAAGAGATA 3389
1015 euAsnYsSerYsgLUleuGInGlyGInGlyLeuYsArgAsn 1031
3390 TGAGACTCTAGACAGATATGATACACAGATACAGAGAAAGCTGG 3439
1032 IleAspAsnLeuYsTYrArGlySHrLysAlaAspValGluGlnLe 1048
3440 CTACAGATAGACTTACTTAAAGAAAGAAATGAGAACTAAAGAGAT 3489
1048 uThrArgAspIleGluSerLeuGInGlyLeuYsSerIleGlySerL 1065
3490 TGA.....GAGAAAGAAACACATTTGAGAGAAAGGTCTCAA 3530
1065 euSerAlaIleGluAlaAspLeuYsArgHIsSerGInGlyYs..... 1079
3531 TGCAGGTTTGCAGAAATGAAAGTGAACATCAGAGTTGGAAGAGACATA 3580
1080 GluArgLeuAsnSerGluPheAsnArgTPGInGlyThrLeuSerValTy 1096
3581 GACAAATATATAAAGAAATCATATATTGGCATTAGGCGACAGAAAGCTTA 3630
1096 rGlnSerAnIleSerYsHIsLysGInGluLeuYsLeuSerGInTyL 1113
3631 TGAAGAGAAATATTTATCTTAAAGAAAGACTTCGAGAACCAATTC 3680
1113 YAspIleGluYsArgTYrThrAsnGlnPheLeuGInLeuYsTYrThr 1129
3681 GGGATGCTGAGAAAGATAGAGAAATGATGATTGTATGAGCAGACAA 3730
1130 GluMetAlaAsnYsAspLeuAspArgTYrTYrThrAlaLeuAspYsAl 1146
3731 GAACCTGTGAACAGATCTGATATTATTATTAAGACTCTTGACCAACG 3780
1146 AleuMetArgPheHIsSerMetLysMetGluGluIleAsnYsIleIleL 1163
3781 AATATATGAATTCACAGATGATAAAGAGAAATCAATTAATTAATAC 3830
1163 YsGluLeuTYrPGLnGInTYrArgGlyGlnAspIleAspTYrIleSer 1179
3831 GTGACCTGTGGGAGATACCTATCTGTGACAGATATTCAATATACATAGA 3880
1180 IleAsnSerAsp.....SerGluGlyAlaGlyThrArgse 1191
3881 ATACGCTGATGCGGATGCAAAATGATACGCTTGTGATTAAGAGGGA 3930
1191 rTYrSerTYrArgValAlMetGlnThrYsAspAlaGluLeuGluMetA 1208
3931 TTATATACACACAGTGTGATCTGAAAGGAGACACAGCTTGATGATATC 3980
1208 rGGLYArCySerAlaGlyGlnYsValLeuAlaSerLeuIleIleArg 1224
3981 GAGGACGATGCTGCTGACAAAGGATTTAGCTTCACTCATTCATTCGC 4030
1225 LeuAlaLeuAlaGluThrPheCyLeuAsnCyGlyIleLeuAlaLeuAs 1241
4031 CTGGCCTGGCTGAAACGTTTGTCTCAACCTGGCATCATTCCTTGA 4080
1241 PGLuProThrThrAsnLeuAspGlyProAsnAlaGluSerLeuAlaAla 1258
4081 TGAGCTACACAAATCTTGACCGAGAAACATTCATCTTCCACATG 4130
1258 lAleuLeuArgIleMetGluAlaArgLysGlyGlnGluAsnPheGlnLeu 1274
|||||.....|

```

```

4131 CTCTGCTGAGATATATAAAGTCGCTCACAGCAGCTAAGCTCAGCTT 4180
1276 lIleValIleThrHIsAspGluArgPheAlaHIsLeuIleGlyGlnArgL 1291
4181 CTGTATATCATTATGATGAAATTTTGTGACCTTTTGTGAGCTCTCA 4230
1291 nLeuAlaGluYsTYrTYrArgValSerYsAspGluAsnGlnHIsSerI 1308
4231 ATATGTGAGAAATTTCTACGATTAATAAGAACATGATCAGTCTCAG 4280
1308 lIle 1309
4281 AGAT 4285

```

seq.name: /SIDS1/gcdata/geneseq/geneseq/NA2000.DAT:AAA09326

seq.document\_block:  
ID AAA09326 standard; DNA: 6773 BP.

```

AC AAA09326;
DE 10-AUG-2000 (first entry)
DE Human cancer associated antigen precursor DNA, clone NY-REN-53.
KM renal cancer; cancer associated antigen precursor; diagnosis;
KM cytostatic; KIAA0336 mRNA tag; ss.
OS Homo sapiens.
PN W0200020587-A2.
PD 13-APR-2000.
PF 04-OCT-1999; 99WO-US22873.
PR 05-OCT-1998; 98US-0166300.
PR 05-OCT-1998; 98US-0166350.
XX (LUDW-) LUDWIG INST CANCER RES.
XX PA
XX PI Obata Y, Golt I, Turci O, Sahlin U, Pfreundschuh M, Scanlan MJ;
XX PI Stockert E, Chen Y, Old LJ, Jager E, Knuth A;
XX DR WPI; 2000-303774/26.
XX PT Preventing, diagnosing and/or treating disorders associated with
XX PT abnormal expression of human cancer associated antigens
XX PS Claim 57, Page 97-99; 121pp; English.
XX XX
XX CC AAA09321-45 were isolated by SEREX screening from a renal cancer
XX CC cell line 1973/10.4. Homology searching revealed that these clones
XX CC correspond to known genes. The present sequence has identity with the
XX CC KIAA0336 mRNA tag gene. The genes encode cancer associated antigen
XX CC precursors. These gene products are useful in methods for preventing,
XX CC diagnosing and/or treating disorders, especially cancer, associated with
XX CC abnormal expression of human cancer associated antigens. The method
XX CC comprises contacting a sample from a subject with an agent that
XX CC specifically binds to the nucleic acid molecule or expression product
XX CC (or fragment) complexed with a human leukocyte antigen (HLA) molecule
XX CC and determining the interaction between the agent and the nucleic acid
XX CC molecule or the expression product as a determination of the disorder.
XX XX
SQ Sequence 6773 BP; 2566 A; 1092 C; 1363 G; 1752 T; 0 other;

```

#### alignment\_scores:

Quality: 478.50 Length: 1249  
Ratio: 0.700 Gaps: 59  
Percent Similarity: 54.764 Percent Identity: 21.057

alignment\_block:



US-09-538-396-2 x AAA09326

Align seg 1/1 to: AAA09326 from: 1 to: 6773

```
91 LysAspValValCysIleArgSerPheGlnLeuThrGlnLysAlaSerLy 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 AAAGAGTGTGATGTC.....CAGATGAGACATCAGCTAAGAAACATGA 750
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 smetGluPheLysAlaIleGlnSerValLeuGlnThrIle.....Asp 122
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
751 AGCAGAGTAATAATAGTGAACGAGCTAAAGAACTAAGTAAACAAAT 800
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 RohsThrgLysLys...ValCysLeuSerTyrArgCys..... 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 GTGAGGCAAGTGAAGAAACATCCAGAAAGAAATGAAATGAGACTTGA 850
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 .....AlaAspMetAspArgLysIleProAl 143
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
851 AATTAAAGAAAGCCACCTCAATGCAACCAACCAACATCAGATATGTC 900
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 AluMetGlyValSerLysAlaValLeuGlnAsnValIlePheValHisG 160
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 TATCTCTTGCAGAAATAATACATTGTGTGAACAAAGTAGTA..... 940
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 LnaPepLysSerAsnTyrProLeuGlnAspProSerThrLeuLysLys 176
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
941 ..AMTGAAGAAAGTCMAACACTTAGAAGAT.....ACCTTAAAGAA... 979
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
177 PheAspArgIlePheSerAlaThrArgTyrThrLysAlaLeuGluVal 193
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
980 .....CTTGAATCTCA 990
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 eluLysLysLeuHisLysAspGlnMetGlnLysIleLysThrPheArgLeu 210
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
991 ACACAGTATCTTAAAGATGAGTAACTTATATGATTAATCTTAAAGTAA 1040
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 yslLeuGln...AsnLeuGlnThrValLysAspGlnAlaHisLysLeu 225
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1041 AACTGGAATGATGCTCAACATTAAGAGATGTTTTCATGAAACGCG 1090
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 GluAsn.....IleAlaGlnAspGln 233
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1091 GAAAGCTTAAGATTTAAATTAATTAATTAATTAATTAATTAATTAAC 1140
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
233 ulysSer.....AspAlaSerLysSerGlnMetGlnLysLeuLysGlu 248
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1141 GGGCTGTGTAATTTGAAAAATTTAAATCTGAGCTACAGCTTAAATTAAC 1190
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 yslIleCysGlyThrGlnArgGluIleLeuGlnMetGlnThrSerLeuAsp 264
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1191 ACTTTTGTCTATCTGTAGAACAGCATTAACAGAAAGTACAGAGCTTTAAG 1240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 GluLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSe 281
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1241 GAA.....CAACATCAAAAAAGAAATATCAGAACTAAATAGAGAC 1278
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 rThrLeuLeuThrGlnGlnHisGlnLysLeuAla..... 292
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1279 AATTTTGTCTAGATTCAGAAAAAGAAAAATTAACATTAACTTTGAATAC 1328
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 ..AlaLeuSerGlnLysGlnAsnGlnLysPheArgLysLeuMetGlnTyr 308
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1329 AGGGCTTAAGAACAGCTGTGAACCTACAGCAAGAAAGCAACAAAGCA 1378
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 GlnThrLysPheGlu.....GluArgIleAlaLeuLeuGlnThrLys 322
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1379 AATTAAATTAATAGAGCTTACGAGAGATTAAGAAATTTTCAAAACAGA 1428
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 sIleSerLysLeuValArgAspMetAspArgLysAlaSerLysSerTyr 339
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1429 ACTGGGGGAATCTGCTGCAAAATAATAGTCMAAG.....TTCGATCA 1472
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 AlaLeuSerLysGlnAsnSerGlnLeuThrHisGln..... 350
```

```
1473 TGAAGCAACAGACAGCATCTGATGTT...CATGAATCGACAGACAGCTC 1519
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 .....IleGlyLysIle 354
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1520 AGAAGCTCTTTTACTGAAAAGATGCCCTTCTGAAACTGTGAATGCCCT 1569
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 uGlnAlaGlnAlaAspAlaHisLeuThrMetLysHis.....Glu 368
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1570 CCAGGAGAAAATGAAAAGTTACTATCTCAACAGAAATTTGGTACCAGAAC 1619
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
368 rgaSerAspIleLysAsnIleCysThrLysHisAsnLeuGlyProVal 384
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1620 TTGAATATACCATTAAGAACCTTCAGAAAAAGAAATGAGTAGTA..... 1660
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
385 ProGlnHisProPheThrAsnAspValAlaMetAsnLeuThrAsnArg... 400
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1661 .....TACTTACTTGTCTCAGTCAAGCA 1686
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 .....IleLysAlaArgLeuSerSerLeu.....G 409
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1687 TACCATGTTAAAGAAATTAAGAGAAAGATTAATCTTCTTACTGAGGAA 1736
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
409 LnaAsnAspLeuLeuAspLysLysSerAsnGlnAspGlnLeuAspVal 425
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1737 AAGATGATTTTATTAATTAATTAAGTAAATTCATGAAGAAATGATAT 1786
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
426 LeuThrLysHis.....TyrLeuLysIleAsnAl 435
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1787 TTCATTAAGAAATGTGAAGGAAAGAAAGATTTGATTTCTGACTTGGGA 1836
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 aArgTyrSerGluVal.....AspGlyGlnIleGlnSerLysIleG 449
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1837 GAAAGTACAGCAACAAATCCAGTACACAGTCACTAGAACAAAGGTA 1886
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
449 LysSerMetSerGlyIleLeuArgArgArgLysAspLysGlnLysGluArg 465
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1887 ATGATTAACAGAGAGACTAGAGAGACTTA...AAAGAAAGAGATCA 1933
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
466 AspAlaAlaGluValGluLeuSerLysPheAsnLeuSerArgIleAsp 482
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1934 AATGACCAAAAGCTAGAAAACCTATGCTTCAATGAAGATCTCTCGA 1983
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
482 uArgGluArgHisMetGlnIleGluVal..... 491
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1984 AGACAAAGAGATATGTACGCTGAAGTAAAGTCTTTATAGCAAAACA 2033
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
492 .....GluArgLysThrLeuAlaLeuGlyGluArgAspTyr 503
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2034 ATTAACCTGAGTTCAGAAAAAAACAGTTGAGT.....AGGATTTG 2074
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
504 AspSerIleIleSerGlnLys.....ArgThrGlnValTyr 515
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2075 GAGGTTTTGTCTCCAAAAGAAAGATGTATCTTAAAGAACATATAC 2124
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
515 rSerLeuGlnGlnLysIleLysValLeuLeuArgGlnLysAspIleAla 532
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2125 TCAATTAAGAAAAGAACTTCAGTTAATGTGTCAAGACAGATATTTAA 2174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
532 snArgAsnAlaAsp...GluArgVal...LysLeuGlyLeuLysLysAsp 546
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2175 ATTAACCTGTTGAATAATGAGCAAGTTCAGAAAGTATTTGTTAAACTCG 2224
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
547 AlaLeuGlnSerSerLysAspLysLysGlnLysLeuAla.....Asp 561
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2225 TTGTATGTTTCTTAAAGAAATGGGATCAGAAAGTTTCAGAAAGACAGTA 2274
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 uHisLysAspLysIleLysLysValLeuArgGlyArgAsnProPheGln 578
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2275 AGAGAAAGATGTTGTTAATGTCTTACAGGACGCTGTAATCTTGGCAA 2324
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
578 yAspMetLysLysGlnLysGlnAlaPheTyrProValAspLysGlu 594
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

2325 AATAAATGAGAAAAATGCAACTGGCTTT ..... 2356  
595 TYRANGLULEUARGSERGLUAGLUAGLUAGLULEUYSPh 611  
2357 ..... CAGCGTGAATGAAAAAGTATTAGATTAGAAAGAGATTAAAGTG 2400  
611 eThInserLysValThrAspAlaArgGLUInLeuThrLysLeuArg 628  
2401 CCTTCAGAGAGAGAGTGTAGTCTCAGTGCAGAACACTTAAGTCTTTTTC 2450  
628 rGAspMetAspAlaLysArgArgPheLeuAspSerLysLeuGlnSerIle 644  
2451 GAGACTATGAGCAAGAAAGTCTCTTAAGCAAGAGTTAGAA ..... 2494  
645 LeuGlnIleSerAlaAsnValAspMetPheProLysValLeuGlnAspAl 661  
2495 ..... GAAATTCAGCTCAGAAAAAGCCCTGCAGCTGATCTCTAGAA... 2539  
661 aMetAsnLysArgAspGLUInLysArgLeuGlnAsnPheAlaAsnGLYm 678  
2540 .ATGAAGATGCTATGAAAAACAAGCTTGAAT ..... 2575  
678 eTArgGLUmetLeuAlaProPheGlnHisLeuAlaArg ..... Lys 691  
2576 .CAGAACTCTTTAATTCAGTTGAAGAGATCTCAACATGTAAGCAAA 2623  
692 AsnHisValLysProCysCysGLUArgAlaPheThrProAspGLUAs 708  
2624 AGTGAATCCATTAATGAAAAAGAAATGTTTATA ..... AACGAAACA 2667  
708 pGluPheValLysLysGlnArgMetGlnAsnSerSerThrAlaGluArgS 725  
2668 TGAAAACCTTAAGCCACTACTAGAACAAAAGAAATTACGAGTACGAGAG 2717  
725 eTYSAlaLeuAlaMetGlnSerSerAsnAlaGluAlaLeuPheGlnGln 741  
2718 CAGAGTGTACTATTAAAGATCTCTTACGAAATACACCTCTGTAAAA 2767  
742 LeuAspLysLeuArgThrIleTyrAspAlaTyrValLysLeuValGlu 758  
2768 AATGATCCTCTGCTTCA ..... GTAAAGAGCTTGAAAGA 2802  
758 uThrIleProLeuAlaGluLysAsnLeuAsnGlnHisLeuAlaAspGluS 775  
2803 AAAAAATA ..... GAAATCTGGAAGAAAGAAATGCAAGAAAGG 2840  
775 eTGLNYSAlaGlnAlaPheAspAspLeuGluValLeuAlaHisVal 791  
2841 AGAGAAATAATAATAGATA... AATTAAGTTGCCGTAAGCAAGAAAGA 2887  
792 GlnMetAspArgAspAlaValGluAlaLeuLeuGlnProThrAspThrI 808  
2888 GAACCTAGTCTCAGCAAGAAAGAG ..... ACCCAAGACTGT 2922  
808 eAspArgHisValHisGlnIleGlnGlnLeuValLysGluValGluAsp 825  
2923 GAAG ..... GAGGAACTGGAATCTCTGCATCAGCAAGAAAGGCCGT 2963  
825 euGluTyrAlaLeuAspSerSerGlyArgGlyValLysSerLeuGluGlu 841  
2964 TATGTGCTTCATGAGAGATCTCATTCAGAGAGCAAGAAAGCTATTAAAGAT 3013  
842 ILeuGlnLeuGluLeuAsnPheLeuGlnArgThrArgAspThrLeuIleVal 858  
3014 CTTTATTATAGAA ..... TATGAAAGCAGTCCAGCAAGCTGAGAGT 3054  
858 IGLUValAspAspLeuArgAsp... GlnHisArgMetLeuAsnGluAsp 874  
3055 GGAAGAAAGAACGTCTAATAATTGTCAGCATCGTAAT ..... GAGAGCC 3098  
874 eT ..... SerSerAlaGlnValArgThrPHisAsnAla 884  
3099 TTACAGACATTAAGAAATTGCACTTGCAGTGTGAAGCAATAAATCT 3148  
885 ArgGLUgluLysValLysAlaSerSerIleLeuGluArgPheGlnLys 901  
3149 GATTAATGAAGATCTCCTGCTGCT ..... ATGAGACATTTACAGCTTAA 3192  
901 rGluGlu ..... GluLeuValLeuLeuAlaGluGluLysGluGlnLeuI 916  
3193 TGCCAAATATTATAGAGTACAGATTATTAAAGATCCAGAGAACCAAGCA 3242  
916 LeuValGluLysLysLeuLeuGluGluSerLeuAspProLeuSerLysGlu 932  
3243 TGTATACAAAGAAATTAAGAGCTGAAAACTT ..... CAG 3277  
933 LysGlnSerLeuLeuGlnGlnLysArgAlaLeuLysGlnLysLeuAspAl 949  
3278 AAGAACAGAGATTAAGAACATGCGACTACTGTAAAGAACTGAAAGA 3327  
949 uGluTyrHisGlnLeuAlaGluArgLysArgGluPheGln ..... G 963  
3328 ACTTCAGTACAACTTCAAAAGAAAGAAACAGCTTCAGAAACCATGC 3377  
963 InGluLeuAspAlaLeuGluArg ..... LeuAsn 972  
3378 AAGAAATTAGAGCTGTTAAAGAGATGCCCAACAAACCATTTGATGAT 3427  
973 MetLysIleLysGlyTyr ..... LeuAspSerLys... 982  
3428 ATGGAATAGCTGATATGACGTTTGAATGAAGAACTTAATCAAAAGTT 3477  
983 ..... LysAsnGluLysLeuLeuGluLeuGlnLysArgHisValLeu 997  
3478 AACTATTAAACAAACAGATGAAGATTGGAG ..... 3511  
997 YHisSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgIleSerAla 1013  
3512 ..... CAAGAAATTAATAATTCAAAACAGAAACAAAGAAACCTTACAGAA 3556  
1014 GluLeuAsnLysSerLysGluLeuGlnGlnLysGlnLysGlnLysArg 1030  
3557 GAAATTAAGCTTCACTACAGCTTCACTACACAAATTAAGAAAGAAACAC 3606  
1030 gAsnIleAspAspAsnLeu... LysTyrArgLysThrLysAlaAspValG 1046  
3607 CAATAATCAAGCAATGCTGTGAAGAAACCAAAAGAAAGTGCAGATTCNA 3656  
1046 InGlnLeuThrArgAspIleGlnSerLeuGlnGluArgLeuLeuSerIle 1062  
3657 AGCAAGCAGAAACGTACCTTAATACCTTCAAGCATCTTTAAA ..... 3700  
1063 GlySerLeuSerAlaIleGluAlaAspLeuArgHis ..... 1075  
3701 GGTGAGCTGAGAGCAAGCAGCAGCAAGTAAAGTCTTAAATACAGCT 3750  
1076 ..... SerGlnGluLys ..... GluArgLeuAsnSerG 1085  
3751 GCGTGAATTAACATCAGAGAGCACAATAATCCACAGCAGCTGAAACCT 3800  
1085 IuPheAsnArgTrpGlnGlyThrLeuSerValTyrGlnSerAsnIleSer 1101  
3801 CTGCGGAACAGCAGCAGCGTACGCTTAAGTGCATACAGAGAGAGTACA 3850  
1102 LysHisLysGlnGluLeuLysLeuSerGln ..... 1111  
3851 GCACACTACAGAAAGATGCGCTGCTCCAAAGCAGAAACAAGCTACTGTAAC 3900  
1112 ..... TyrLys ..... AspIleGluLysArgT 1119  
3901 CTGTGAATTCAGAGCTACAAAGTCCGAGTCAATAGTTCTTAAACAAAC 3950  
1119 YThrAsnGlnPheLeuGlnLeuLysThrTrpGluMetAlaAsnLysAsp 1135  
3951 AGAAAAATTAATCTATGTCTCAGGCTGAAGTGAAGGCGCTTAACAAAGA 4000

1136 LeuAspArgTyrThrAlaLeuAsp..... 1144  
 4001 AGGAAACATCTGAAATGCTGATTGACCACTAAATCAATTACAGA 4050  
 1145 .....LysAlaLeuMetA 1149  
 4051 TAGCCAAATAACTTACAGTATATGTAACCTCAACATTCGACGT 4100  
 1149 rrpHHisSerMetLysMetGluGluIleAsnLysIleIleLysGlu 1164  
 4101 CTGAACATGATGACCTCTAGAAAGCACAACAAGATGTCGACAGAA 4147  
 seqname: /STD1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV21209

seq documentation\_block:  
 ID AAV21209 standard; DNA; 1664976 BP.

AAV21209;  
 10-NOV-1998 (first entry)  
 Methanococcus jannaschii circular chromosome.  
 Methanococcus jannaschii; methanogenic archaeon; circular chromosome;  
 genome; autotrophic; extrachromosomal element; identification; ds.  
 Methanococcus jannaschii.  
 MO9807830-A2.  
 26-FEB-1998.  
 22-AUG-1997; 97MO-US14900.  
 22-AUG-1996; 96DS-0024428.  
 (GENO-) INST GENOMIC RES.  
 (UNIT) UNIV ILLINOIS FOUND.  
 (UJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 Bult CJ, Smith HO, Venter JC, White OR, Woese CR;  
 WPI; 1998-169145/15.  
 Complete genome sequence of methano-genic archaeon, Methanococcus  
 jannaschii - useful in identification of M. jannaschii genome  
 fragment  
 Claim 13; Page 152-585; 614pp; English.

The present sequence represents the complete 1.66-megabase pair genome  
 sequence of the Methanococcus jannaschii circular chromosome. The  
 present invention describes M. jannaschii open reading frames from the  
 genome sequence. The invention also describes a computer based system  
 for identifying fragments of the M. jannaschii genome that are  
 homologous to target nucleotide sequences, comprising: (a) data storage  
 means comprising the nucleotide sequence of the 1664976, 58407 or 16550  
 bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide  
 sequence at least 99.9% identical to it; (b) search means for comparing a  
 target sequence to the nucleotide sequence of the data storage means to  
 identify a homologous sequence; and (c) retrieval means for obtaining  
 the homologous sequence. The method, which is based on whole genome  
 random sequencing of an autotrophic archaeon M. jannaschii, the genome  
 of which consists of 3 physically distinct elements, a large circular  
 chromosome (the 1664976 bp sequence given in AAV21209), a large circular  
 extra-chromosomal element (the 58407 bp sequence given in AAV21210), and  
 a small circular extra-chromosomal element (the 16550 bp sequence given  
 in AAV21211), can be used in the identification of M. jannaschii genome  
 fragment.  
 Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;

alignment\_scores:  
 Quality: 465.50 Length: 1356  
 Ratio: 0.721 Gaps: 54  
 Percent Similarity: 47.640 Percent Identity: 19.985

alignment\_block:  
 US-09-538-396-2 x AAV21209

Align seg 1/1 to: AAV21209 from: 1 to: 1664976

7 MetLeuIleLysGlyIleArg.....SerPheAspProAspAsnLys 21  
 198975 ATGATTAATAAGAAATAGATGATGATTAAGTCAATGATGATTC 199024  
 21 nValIleThrPheLysProLeuThrIleValGlyProAsnLys 38  
 199025 AAGAAATTAAGTTGAAAAGGATGTTGCAATTAATGAGAGATGAA 199074  
 38 LagLysThrThrIleGluCysLeuLysLeuSerCysThrGlyGlu 54  
 199075 GTGCAAAATCATCTATCTTGAACAGCTGTTCTTCCCTGTTGGC... 199121  
 55 LeuProProAsnSerArgSerGlyHisThrPheValHisAspProLysVal 71  
 199122 .....GCAGCAGTAATTTTAATTAGCACACAAATAT 199153  
 71 LAlaGlyGluThrGluThrLysGlyGln.....IleLysLeuA 84  
 199154 A.....ACCAAGGAAAAAATCCGTTTATGTAATGG 199188  
 84 rrpHHisSerMetLysMetGluPheLysAlaIleGluSerValLe 100  
 199189 ATTTTGA...GTCATGGAACAACACTACAAATTAATCAGAGATATGAT 199235  
 101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerValLe 117  
 199236 TCTGGAAGAGGAGGAGCTAAGCTC...TATAAGATGGAAGCCTTACGC 199282  
 117 uGlnThrIleAsnProHisThrGlyLysValCysLeuSerLysArgC 134  
 199283 TACTACAAATTAAGTCA..... 199298  
 134 YsAlaAspMetAspArgGluIleProAlaLeuMetGlyValSerLysAla 150  
 199299 .....GTTAATAAGCAGTAATGAAATCTTAGGCTTGATAGAAAT 199340  
 151 ValLeuGlnAsnValIlePheValHisGlnAspGlu..... 162  
 199341 ATGTTCTTAACCTCATATATATTAATTAACAGGAGGAGATAGTAATTTT 199390  
 163 SerAsnTrpProLeuGlnAspProSerThrLeuLysLys..... 175  
 199391 GAGTTTAAACCCCTCCGAAAAATGGAACAGTTGCGAAACTTTTGGCA 199440  
 176 .....LysPheAspAspIlePheSer.....AlaThrArg 185  
 199441 TAGATGAGTTGGAATAATGCTATCAAAAAATGGGAGATTTGTAAGCA 199490  
 186 TyrThrLysAlaLeuGluValIle..... 193  
 199491 TATGAAAAAGATTAAGAAAGATTGAAGAGAGTTGATTCAGAAAGAA 199540  
 194 .....LysLysLeuHisLysAspGlnMetGlnIleLysThrPheArgL 209  
 199541 TTATGAAAAAGATTAATAATAAATGAGCCCAATTTGMAAGAA..... 199583  
 209 eulysLeuGlnAsnLeuGluThrValLysAspGlnAlaHisLysLeuArg 225  
 199584 .....AAAAATTAATAATTAATGAAATTAATGATTAATCAATTAATAA 199631  
 226 GluAsnIleAlaGlnAspGlnGluLysSerAspAlaSerLysSerGlnMe 242  
 199632 AAGGAATTT...GAAGATTTGAAAAAATTAATTAATGATGAGAAATTA 199678

242 tGluGlnLeuLysGluLys... IleCysGlyThrGluArgGluLeuLeu 258  
 199679 AAAGTGTGTGTATGAAAAATTCATAAACAACCTTGAAAGAAAAGAGAG 199728  
 258 LmetGluThrSerLeuAspGluLeuArgArgLeuGlnGlyGlnIleAsp 274  
 199729 CTTTAGAGCTTAAAAATCAAGAGCTTAAAAATTTAGATAATGATTAAAT 199778  
 275 IleLysAlaThrGluArgSerThrLeu... LeuThrGlnGlnIleGln 289  
 199779 ACCTGTGTGAGCAAGAGAACTCTAAATAGACATTAAGATGAAATGGA 199828  
 289 uLysLeuAlaIleLeuSerGluGlnGlnAspThrAspGluLeuLeu 306  
 199829 AAATATATAATCATTAAGTGTGATGAGATTAGGAAAAATTGAGAGCAGATTAA 199878  
 306 etGluThrGlnThrLysPheGluGlnArgIleAlaLeuLeuGluThrLys 322  
 199879 GAGATTAAGAGCTCATTAAGATTTATTTAAAA... 199913  
 323 IleSerLysLeuValArgAspMetAspAspGluIleAspTyrSerSerVa 339  
 199913 ..... 199913  
 339 IleuSerLysGlnAsnSerGluLeuThrHisGluIleGlyLysLeuGln 356  
 199914 TTAAACAAGCAGCTTGAGATTAATAAAGAGACATTCGAAAAATGTGAAA. 199961  
 356 IacGluAlaAspAlaHisLeuThrMetLysHisGluArgAspSerAspIle 372  
 199962 ..... GAATTCATCAACAAAGCTAAAGTATAGAGATGATATT 199997  
 373 LysAsnIleCysThrLysHisAsnLeuGlyProValProGluHisProPh 389  
 199998 GATAAATTAGATTA... 200012  
 389 ethAsnAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuS 406  
 200013 ..... CTATTAAATTAATAAATAAAGATGAAATGTG 200040  
 406 erSerLeuGlu... AsnAspLeuAspLysLysSerAsnGlu 420  
 200041 AAAGAGTGAACACTTAAAGATTTGCTTGAAAGACTTAATAAATCTTAAT 200090  
 421 AspGlnLeuAspValLeuThrLysHisIleLysLysIleAsnAlaArgTy 437  
 200091 GAAGAGATTGA... 200102  
 437 rSerGluValAspGlyGlnIleGlnSerLysIleGluSerMetSerGlyI 454  
 200103 ..... AAAATGAAAAAATATAAAGAA 200124  
 454 leLeuAlaArgArgLys... AspLysGluLysGluArgAspAla 467  
 200125 TATGTGAAGAGCTGCAAAAGAACTATGAGAAATATTGAAATTAAGAAGA 200174  
 468 AlaGluValGluLeuSerLysPheAsnLeuSerArgIleAspGluArgG 484  
 200175 AAGCGTGAAGAACTAACTAACTTA... 200204  
 484 uArgHisMetGlnIleGluValGluArgLysThrLeuAlaLeuGlyGluA 501  
 200205 GAGATATATACATTGCTTCAGAGAAAAAATCCATGAAAAAATAATTA 200253  
 501 rGAspTyrAspSerIleIleSerGln... LysArgThr 512  
 200254 ACGATTATGAACAACAGAAATTAATAAATCTTTGAAGAAAAAATAATAT 200303  
 513 GluValLysSerLeuGlnGlnLysIleLysValLeuLeuArgGluLysAs 529  
 200304 GACATTGAAGATTTGAAATTCATTAAAGAGATGAGAGAAAAAAGAA 200353

529 pIleIleAsnArgAsnAlaAspGluArgValLysLeuGlyLeuLysLysA 546  
 200354 AGTCTTGAAAAATCTACAAAAAGAAAGATGAGACTTAACAAAAAAGCTTG 200403  
 546 spAlaLeuGlnSerSerLysAspLysLeuAsnGluIleValAlaGlnHis 562  
 200404 CGGAAATTAACAGTGAATTTAAAGGCTGAAAAAATTTTGAAGAACTT 200453  
 563 LysAspLysIleLysLysValLeuArgGlyLysArgProPheGluLysAs 579  
 200454 AAAGAACTGAGGAAATGCCCCATGTATTAACCCAAATAGTGAATA 200503  
 579 pMetLysLysGlu... IleAsnGlnAlaPheThrProValAspLysGluT 595  
 200504 TAAAAAGTGAATTAATTAATCAACATTAACCAAGCAGTAAACATTAAT 200553  
 595 TyrAsnGluLeuArgSerLysSerGlnGluAlaGluGlnGluLeuLysPhe 611  
 200554 ATACGAAATTCGACGAA... 200570  
 612 ThrGlnSerLysValThrAspAlaArgGluGlnLeuThrLysLeuArgAr 628  
 200571 ATAAATTAATAAATTAAGAGATTGAATAAGATTAAGAAATTAAGAA 200620  
 628 GAspMetAspAlaLysArgArgPheLeuAspSerLysLeuGlnSerIleL 645  
 200621 AGAAATTTAT... AAAGAGAAATCTTAAGACACTAAGACCTAATATC 200667  
 645 euGlnIleSerAlaAsnValAspMetPheProLysValLeuGlnAspAla 661  
 200668 TTGAATAACAAGTCAGATTGAAGAAATTAAGAAATTAATAAATTAAT 200717  
 662 MetAsnLysArgAspGlu... GlnLysArgLeuGlnAsnPheAla... As 676  
 200718 AAAGAGCAGTATGAGATCAATTAATAAATATCCCACTATGTAATTA 200767  
 676 nGlyMetArgLysMetLeuAlaProPheGluHisLeuAlaArgLysAsnH 693  
 200768 CGGA... AACCCAGTGATGAGTA... 200789  
 693 IValLysProCysGluArgAlaPheThrProAspGluGluAspGlu 709  
 200789 ..... 200789  
 710 PheValLysGlnArgMetGlnAsnSerSerThrAlaGluArgSerLys 726  
 200789 ..... 200789  
 726 sAlaLeuAlaMetGluSerSerAsnAlaGluAlaLeuPheGlnGlnLeuA 743  
 200790 ..... TTGAAGACATTAAAGCCAGCTGA 200814  
 743 sPLeuLeuArgThrIleTyrAspAlaTyrValLysLeuValGluGluThr 759  
 200815 ACAAAATTAATAAATCTTATTAACCAATCTATCAAGCTGTAAGTAATTA 200864  
 760 IleProLeuAlaGluLysAsnLeuAsnGlnHisLeuAlaAspGluSerG 776  
 200865 AATAGCTGATGAGAGAAAGT... 200885  
 776 nLysAlaGlnAlaPheAspAspLeuLeuGlyValLeuAlaHisValGlnM 793  
 200885 ..... 200885  
 793 etAspArgAspAlaValGluAlaLeuLeuGlnProThrAspThrIleAsp 809  
 200885 ..... 200885  
 810 ArgHisValHisGluIleGlnGlnLeuValLysGluValGluAspLeuG 826  
 200886 ..... ATTAAGAAATGAATTAAGGAAATGAAAAATATGCT 200920  
 826 uTyrAlaLeuAspSerSerGlyArgGlyValLysSerLeuGlnIleG 843

```

200921 AAGTGGATGGATAAA.....GAAAAATGTA 200946
      :: :::::
      InLeuGlueuAnpheLeuGlnArgThrArgSPThreuLeuValGlu 859
      :: ||||| :::::
      200947 GAGAGAGCTGAACAATATAGAGAGATGAGAGA.....GAA 200984
      :: ||||| :::::
      860 ValAspAspLeuArgAspGlnHisArgMetLeuAsnGlu.....AspMe 874
      :: ||||| :::::
      200985 ATAAACGATTAAGACACAAA.....TTAAATGAACTTAAATAA 201025
      :: ||||| :::::
      874 tSerSerAlaGlnValArgThrHisAsnAlaArgGlnGluValLysa 891
      :: ||||| :::::
      201026 GGAATAAGAACTTAATAGAAATGAGATAGAGAGCTCTTAAAG..... 201068
      :: ||||| :::::
      891 LaserSerIleLeuGlnArgPheGlnLysSerGlnGlnGluValLeu 907
      :: ||||| :::::
      201069 .....TTTGATTAATATATAGGAATATTAGGT... 201095
      :: ||||| :::::
      908 LeuAlaGlnGluLysGlnLeuIleValGlnLysLeuLeuGlnGlu 924
      :: ||||| :::::
      201096 .....CTAACTGA 201103
      :: ||||| :::::
      924 uSerLeuAspProLeuSerLysGlnLysGlnSerLeuLeuGlnGluTyr 941
      :: ||||| :::::
      201104 AAATATTGAAGAGCTTAAATAATTTAAAGATGGCTTGAAGAGATTATA 201153
      :: ||||| :::::
      941 snAlaLeuLysGlnLysLeuAspGlnGluTyrHisGlnLeuAlaGluArg 957
      :: ||||| :::::
      201154 ATATATGCCAACTCTAAGATT.....TTAGCAATATGATTAACATT 201191
      :: ||||| :::::
      958 LysArgGlnPheGlnGlnGlnLeuAspAlaLeuGlnLysArgLeuAsnMetLys 974
      :: ||||| :::::
      201192 AAGAGGAAGTATATAAGAA.....GA 201214
      :: ||||| :::::
      974 sIleLeuGlnLysLeuAspSerLys.....LysAsnGlnLysLeuL 988
      :: ||||| :::::
      201215 TATTTGAATTTACTTAACACAAATTTTAGAGGTTAATTAAGCAATTA 201264
      :: ||||| :::::
      988 ysgIleuGlnGlnLysArgHisValLeuCysHisSerGlnLeuGlnSerCys 1004
      :: ||||| :::::
      201265 ATGATATATAGAAAGAAATATCTACATTAACCAAAACTTGATGAATATA 201314
      :: ||||| :::::
      1005 MetAlaLysGlnGlnArgIleSerAlaGlnLeuAsnLysSerLysGlnLeu 1021
      :: ||||| :::::
      201315 AACATCAATGAGAGAA.....GAACATTAATAAAAGAGCT 201352
      :: ||||| :::::
      1021 uLeuGlnGlnGlnGlnGlnLeuLysArgAsnIleAspAspAsnLeuLysT 1038
      :: ||||| :::::
      201353 CTATGAAAT.....AAGAGACAGAACTTGATTAACGTAAAGAG 201390
      :: ||||| :::::
      1038 yTArgLysThrLysAlaSP.....ValGlnGlnLeuThrArgAspIle 1052
      :: ||||| :::::
      201391 AACAAAAACGAAATATGAGCAGCAATTTGAATTTAAAAAAGATGTT 201440
      :: ||||| :::::
      1053 GluSerLeuGlnGlnArgLeuLeuSerIleGlySerLeuSerAlaIleGln 1069
      :: ||||| :::::
      201441 GAAAGTTTAAACCAAGATTA..... 201461
      :: ||||| :::::
      1069 uAlaAspLeuLysArgHisSerGlnGlnLysGlnArgLeuAsnSerGlnP 1086
      :: ||||| :::::
      201462 .....AAAGAA..... 201467
      :: ||||| :::::
      1086 heaAnaTGTGTGlnGlnLysThrLeuSerValTyrGlnSerAsnIleSerLys 1102
      :: ||||| :::::
      201468 .....ATGCTAATTTAGAGAAA 201485
      :: ||||| :::::
      1103 HisLysGlnGlnLeu.....LysLeuSerGlnTyrLysAspIleGlnLysArg 1118
      :: ||||| :::::
      201486 GAAAAAGAGAAATTAACGAACTTGTTCATATATTAGCAAGGTTAGAG 201535
      :: ||||| :::::
      1118 gTyrThrAsnGlnPheLeuGlnLeuLysThrThrGlnMetAlaAsnLysa 1135
      :: ||||| :::::

```

```

201536 AATATTGGTAGA..... 201548
      :: ||||| :::::
      1135 spLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPheHis 1151
      :: ||||| :::::
      201549 .....AATGATTTTCA 201560
      :: ||||| :::::
      1152 SerMetLysMetGlnGlnIleAsnLysIleIleLysGlnLeuTyrPngln 1168
      :: ||||| :::::
      201561 GCATATTTAAGACAAAATATGTTCCATTATATCCAAAAAATTTAAATGA 201610
      :: ||||| :::::
      1168 nThrTyrArgGlnLysAspIleAspTyrIleSerIleAsnSerAspSerG 1185
      :: ||||| :::::
      201611 AGCATTTAGAGATTGACTTGCTTATAGCTTTTGAAGATC..... 201653
      :: ||||| :::::
      1185 LuGlnAlaGlnLysThrArgSerTyrSerTyrArgValValMetGlnThrGly 1201
      :: ||||| :::::
      201654 .....ACTAAAGATTTGAAGTAAAGATTCATGCTCCAAATGA 201692
      :: ||||| :::::
      1202 AspAlaGlnLeuGlnMetArgGlyArgCysSerAlaGlnLysValle 1218
      :: ||||| :::::
      201693 GTTTTAACCTTGAC.....AATTTAAGCGGTGAGACAGAGATAGC 201733
      :: ||||| :::::
      1218 uAlaSerLeuIleIleArgLeuAlaLeuAlaGlnThrPheCysLeuAsn. 1234
      :: ||||| :::::
      201734 GGTAGCTCTCTTAAATAGATTAAGCATTAAGCTTTAATTGGAATATA 201783
      :: ||||| :::::
      1235 .....CysGlyIleLeuAlaLeuAspGlnProThrThrAsnLeuAsp 1248
      :: ||||| :::::
      201784 GGGTTGAGTGC.....ATTATATTGATGATGACCACTGTATATTAGAT 201827
      :: ||||| :::::
      1249 GlyProAsnAlaGlnSerLeuAlaAlaLeuLeuArgIleMetGlnAl 1265
      :: ||||| :::::
      201828 .....GAAATTAAGAGGCAAACTTACTAATCTTT... 201860
      :: ||||| :::::
      1265 ArgLysGlnGlnGlnLysAsnPheGlnLeuIleValIleThrHisAspGln 1282
      :: ||||| :::::
      201861 AGGAGGTTAAGAGCATTCACAGATGATTAATTAACCCACACAGAG 201909
      :: ||||| :::::
      1282 rgrPheAlaHisLeuIleGlnArgGlnLeuAlaGlnLysTyrTyrArg 1298
      :: ||||| :::::
      201910 AGCTT.....GAAGATGTAGCAGATGTGATTAATCAAT 201941
      :: ||||| :::::
      1299 ValSerLysAspGlnAsn 1304
      :: ||||| :::::
      201942 GTTAAAAAAGATGGAGAT 201959
      :: ||||| :::::

```

seq\_name: /SID1/gcgcdata/geneseq/geneseq/NA1996.DAT.AAT58751

seq\_documentation\_block:

ID AAT58751 standard; CDNA; 4779 BP.

XX AAT58751;

AC

XX

DT 09-MAR-1997 (first entry)

XX

DE Nucleolar/endosomal auto-antigen.

XX

KW Auto-antibody; p162; rheumatic disease; antigen; diagnosis;

KW gene therapy; ss.

XX

OS Homo sapiens.

XX

EH Key

FT CDS

FT Location/Qualifiers

FT 71..4306

FT /tag= a

XX

PN DE19515514-C1.

XX

PD 12-SEP-1996.

XX

PF 27-APR-1995; 95DE-1015514.

XX

PR 27-APR-1995; 95DE-1015514.



579 pmetleyysglluleansgnlalphetipro. .... 590  
1303 actccaaagtgaaattatcatattacatagcaaaactcttgagacagacg 1352  
590 ..... 590  
1353 gccacactaggggaagctcatgtaggctgaagaaagacagacagctttca 1402  
591 ..... valaspyglutyrasn. .... gluleuargserly 601  
1403 agtgaamaagttgatgataaagaacaacagctgctattttacacatcaa 1452  
601 sserlndlnlualuglndlnlueuluyrphethrlnserlyvalpha 618  
1453 actttctcgtggtaagagcagcttgaaagaaaaagttacaaattctacag 1502  
618 apalarglulndlnueulhrys ..... 625  
1503 aatttcagacatcatattaatatmaaaacaaacagatccaaagaaacag 1552  
626 .. leuargargaspmetaspalalysatgtrgphelueaspsertlyle 641  
1553 gctcttcagcaaaagcaccagcgcaaaacttcgaaagctcagatgatatt 1602  
641 uclnserlileuglnlleserialaservalaspmethpherolysall 658  
1603 ggaagaaagttctacactcaaatgtgggataaggaacaaagatccagaac 1652  
658 euglaspalmethasnlaysatrasp ..... gluln 668  
1653 ttgaagcctttattacagaaagctaaagaaattattctattctgaaaaa 1702  
669 lysatgleugluasnphelaasn. .... glyetarglmetle 682  
1703 gaaaggaagatctttanagcaaaaatttcagcctgctgaagagag... ac 1749  
682 ualarpheglunhlsleualabgulyasahis. .... y 694  
1750 tgcctttcttaacaggttacaaagaaaaaaccaatcacactacagagacag 1799  
694 alcysertocysgslunrgralaphethrproasrglulaspgluphe 710  
1800 taactacactaaacagaaagctgaagatcagctcgaagatcattaaaca 1849  
711 vallylsy. .... glnarwetglinaasensertth 721  
1850 gccacagaaatatttcgncgacacagctacaaagcagaaagcacaattctag 1899  
721 ralaagluarsetlysalaaleualmetcglusertersanlaadualal 738  
1900 agctgcacaaagccgncctttcccttgaaacatgctccatga... t 1946  
738 euphedlnlndlnueasrylsleuag. .... thlletyaspala 751  
1947 taatagtcaatttaattatgaagcaagaaagctctccacagcttgacata 1996  
752 tyuvallylsleuvalgluculuthrleferoleualaglu. .... 764  
1997 cagattmaaaacaaaacsgaataattactatcagcagaagaaacgaaaaac 2046  
765 ..... lyasnlueasnglnhlsleualasrgluserlndlnyua 778  
2047 tgcctcaaaagcagatcttcacaaatcatttg. .... gacacagctcaa... 2089  
778 laadlnlarheasryrleuouglvalylealalnlsvalglmetkar 794  
2090 .. aatgcattaccagatgataaagcagcagaaattmaataagatttactag 2137  
795 argaspalavalglulaleuouglndlnprothrasertthleasargh1 811  
2138 ttggatcaggtacstgcgaagattacaaagcaag ..... caagaca 2178

811	gvalhnglulleglningluleuvalysglulvalglasp.....	824
2179	ttgcagtcagctgggaaatctattttaaagattatnaagagaaaactct	2228
825	..leuglultrylaleuaspserseglaryglvalylvsseleuglu	840
2229	cttttggaaacgaaaaacgaagagctagaaagctcaaatatgaagaaactgaaa	2278
841	glullleugluleu.....leuasnphlele	849
2279	gctgattagcttgaaagtttaagcaagcagagcagcctttgcagaatct	2328
849	ucllnatrgtlaragasp.....thleulevalglulvalaspaspl	863
2329	acaaagcamaacgacgctgaaacacagatttgagctcagaccacagat	2378
863	euargaspplnlsarvmetleuasngluaasmetseleleaglval	879
2379	tgagttaacacacttaaatggagaagaaatnratctcagacacaaattg	2428
880	argtrphsasnalarglulglulysvalylsalaserleleugl	896
2429	.....gactctcagaaaaaaatctgagcccttgaaatgacacagca	2469
896	uargphleugllysergluglulgluleuvalleuenaajolululysg	913
2470	aaagcttaccaagcagaaggaagaaaaaaacaaactctgaaacaaatatttg	2519
913	lglulneuilevalglulyslyseleuglulgluserleuasp.....	927
2520	aaacttttaagtcanaagaaacaaagattcagcatgagatttgaatnaacga	2569
928	.....proleuserlysglulysgluse	935
2570	attcaaaacaaagtaacagaaactcaaaaagcagaaaaggaagaaagc	2619
935	rlleuenglultryarasnalauleuylsleuasnplulultryh	952
2620	ttttaaagcagagcttctacagttaaagacacaaactatcaaaaagttctg	2669
952	lsglneuajagluartrylsarvgluupheglngluleuaspalaleu	968
2670	attctttgaaaaaactctaaagatggatttgaagaaagc.....	2707
969	glyargleuasnmetlyslellys..glytryleuasnpserslylsas	984
2708	.....aatcagaaagaaaaacccgctatatattagacttgaaaaaaac	2748
984	nglulysleuylsleuenglultryarhstvalleucyshserglnl	1001
2749	t.....tgcacaaagattta.....aagatctcaactt...caagtcga	2783
1001	euolnsercysemetlalysglnlarygllelelealgluleuasnlys	1017
2784	tggaataacacactttaagaaagaaaggaactgaaaaagctcacttgaaaaa	2833
1018	serlyseglu.....	1020
2834	gaaagagagccttcacatcagttgaaattgcaactcaattcattcagaca	2883
1021	..leuenglulnglulnglulngluleuylsaryasn.....	1031
2884	acaactttaaaccagcccaaaaatcttttaaaacaaatgaaaaaagagac	2933
1032	..lleasparasnleuylslyarglysthrlysalaaspyalglu	1046
2934	aaacacttcaggcggaacata.....aatgacacttaagcaatcaagtgaa	2977
1047	gluleuthrtryaspallleglulserleuengluluartryleuaserleagl	1063
2978	caaaagaaaaaaacaaattgaaagactccaaagagag...ctttaaattatgc	3024
1063	yserleuseralalleglulalaspuleuylsarghshsergln.....	1077

```

3025 TGTATTACAGACAGACAGCTTGAAGATTAACATACACACAGCTTAACAC 3074
1078 .....GluysGluArgLeuAsnSerGluPhe 1086
3075 AGCAGCCCAAGAACTTGCAGCAGAGAAAGAAA..... 3109
1087 AsnArgTrpGlnGlyThrLeuSerValTyrGlnSerxsnLleSerGlySH 1103
3110 .....ATATAGATTAATCAAAACCAATGATGAAAAAG 3141
1103 slysGlnGluLeuLeuSerGlnTyrAspLleGluLysArgTyr 1120
3142 TCAGGA.....ACTTCAAAACGCTCAATGATTTCT 3176
1120 hrasGlnPheLeuGlnLeuLysTrpThrGluMetAlaAsnLysAspLeu 1136
3177 ATGGGAGGAATGTGAATCTCTGACCCAGCAGCAAGATCTTAAAGTCTGA 3226
1137 AspArgTyrTyrThrLysLeuAspLysAlaLeuMetArgPheHisSerMe 1153
3227 GAGAGAAAGCTTTCTGTACGACAGGAGACTGATTC..AACAGAA 3273
1153 llysMetGluGluLysAsnLysLleLysGluLeuTrpGlnGlnThr 1170
3274 TCAAATGTGAATCAAAATTAATGATTCAGAAAGCTG...AAGAGTG 3317
1170 yrarGlyGlnAspLleAspTyrLleSerLleAsnSerAspSerGluGly 1186
3318 CCAGAGCT.....ACATGGAGCAGGATTC..... 3343
1187 AlaGlyThrArgSerYrSerYrArgValValMetClnhArgLysAspAl 1203
3344 .....GCAAGAA 3351
1203 aglLeuGluMetArgGlyArgYsSerAlaGlyLysValLeuAsn 1220
3352 AGAAGACGCAATTCAGAGAGCGATGAAGCATACAGACATTCAGAAAG 3400
1220 erLeuLleLe.....ArgLeuAlaLeuAla 1228
3402 AAAAGTCACTGAAGAAAGAAAGAACTGGTAATGAGAGTCTAAATTGGCA 3451
1229 GluThrPheCysLeuAsnYcs.....GlyLeuAlaLeuAs 1241
3452 GAATGATGAAGAAATTAATATGTACACAAAGAAAGAAATCTAAACTTAA 3500
1241 pGlu.....ProThrAsnL 1247
3502 CGAAGAACTCAATCCCAACAACTAGAAAGCATTAAGAGATTAACAATC 3551
1247 euAspLysPro..... 1250
3552 TTAAGATGCTAAACAGACTTCTAATTCAGAGAAATTAGAACTTCAAGA 3600
1251 AsnAlaGluSerLeuAlaAlaAlaLeuLeuArgLleMetGluAlaArg 1267
3602 AAGCGGACATCCCTGAAGCAGCT.....CTGACACGAGGAAGAG 3642
1267 scLysGlnGluAsnPheGlnLeuLleValLleThrHisAspGluArgPhe 1284
3643 AATTCAGCAGATCTTAAGAACCCAGGTGAAGAAAGGAAGAG..... 3685
1284 lnhIsLeuLleGlyGlnArgGlnLeuAlaGluLysTyrTyrArgValSer 1300
3686 .....GAGCTGAAGAAAGATTT.....ATTGAG 3709
1301 LysAspGluAsnGlnHisSerLleLleGluSerGlnLeu 1314
3710 AAGGAAGCTAAGTGCATCTCCGAATTAAGAAAGAAAGAGTA 3751

```

```

seq_documentation_block:
ID AAX26819 standard; DNA: 9626 BP.
AC
AA26819;
AT
22-JUN-1999 (first entry)
DE
Nucleotide sequence of centromere-associated protein-E (CENP-E).
KW
CENP-E; centromere-associated protein-E; ATPase activity;
KW
plus end-directed microtubule motor activity; chromosome congression;
KW
microtubule binding activity; chromosome movement; mitosis;
KW
cell proliferation; tumor; metastasis; vascular malformation;
KW
inflammatory disease; immune disease; angiogenesis; hypertension;
KW
restenosis; fungal infection; selective herbicide; fungicide;
KW
insecticide; plant growth regulator; activator; cancer cell marker; ss.
XX
OS
Xenopus sp.
XX
PN
W09913061-A1.
XX
PD
18-MAR-1999.
XX
PE
10-SEP-1998; 98WC-U519231.
XX
PR
11-SEP-1997; 97US-0058645.
XX
PA
(REGC ) UNIV CALIFORNIA.
XX
PI
Cleveland DM, Goldstein LSB, Sakowicz R, Wood KM:
XX
WPI: 1999-229233/19.
XX
DR
P-PSDB; AAY01632.
XX
PT
Centromere-associated protein-E and related nucleic acid
XX
PS
Claim 13; Page 67-73; 77pp: English.
XX
CC
The present sequence encodes CENP-E (centromere-associated protein-E)
CC
of Xenopus. The protein has at least one of plus end-directed microtubul
CC
motor activity, ATPase (adenosine triphosphatase) activity and
CC
microtubule binding activity. CENP-E is the motor that powers chromosome
CC
movement toward microtubule plus ends and is essential for congression
CC
of chromosomes during mitosis. Modulators of CENP-E can thus control
CC
cell proliferation. Agents that modulate CENP-E activity are lead
CC
therapeutic, biocultural and diagnostic agents, e.g. for treatment
CC
of unwanted cell proliferation (typical of many examples are tumors and
CC
metastases; vascular malfunction; inflammatory and immune diseases;
CC
angiogenesis; hypertension; restenosis; and fungal infections), also as
CC
plant-protection agents (selective herbicides, fungicides and
CC
insecticides) and plant growth regulators or activators for improving
CC
yields. CENP-E is also a diagnostic marker for dividing cells, including
CC
cancer cells.
XX
SQ
Sequence 9626 BP; 3540 A; 1692 C; 2104 G; 2250 T; 0 other:
XX

alignment_scores:
Quality: 449.50 Length: 1506
Ratio: 0.634 Gaps: 62
Percent Similarity: 47.078 Percent Identity: 20.053

alignment_block:
US-09-538-396-2 x AAX26819 ..
Align seg 1/1 to: AAX26819 from: 1 to: 9626
100 GlnLeuThGlnLysAlaSerLysMetGlnPheLysAlaAllegLusSerVa 116
||||| :||||:||||:||||| ||| :||||:||||:
2051 GAATTCAGAGAAAAGGTTCTCGATTCGA...AAGCAGCTTGAGAAAGC 2997
116 lLeuglnPrrlLeaenPrroHstThGlyGlnLysAlaLysLusSer...T 132
||||:||||| ||| :||||:||||:||||| |||

```



2998 ATTGAGAAAAAATGCTTGGAGATGAAGTACTGCTATCAGAAAT 3047  
132 yTAcGcysalaAsermetaspArgluleProAlaleuMetGly...Val 147  
13048 ACAAAATTT...TTGCCAAATGAAGTGAATGCTTGAATAAATCAGATC 3091  
148 SerlyAlaValLeuGlu... 153  
3092 AGCAAGGCTTCTGAAGAGATTAATGTATTAAGCAAGAGAGAACTTC 3141  
154 .....AsnValIlePheValHisGlnAspGlu 163  
3142 TGCATCTATTATTAAGCAAAACAAAGATTAATCATCATCAGAGCAATCTGAGC 3191  
163 eTAsnTProLeuGlnAspProSerThr... 172  
3192 AGATTTTAAACATGCTGAGCAAGTGAACACACACAGTCAAAAGTGCAG 3241  
173 .....LeuLysLysPheAspArgIlePhe 181  
3242 CAGACGTAAGCAATTAATCTTGAGATGAAGAAATGCAATGATGATCTTT 3291  
181 eSerAlaThrArgTyr...ThrLysAlaLeuGluValIleL 194  
3292 TGA...AAATATTCAGAAACAAAGAGTGAAGCTGAAGACTTTTAA 3335  
194 yLysLysLeu...LysAspGlnMetGln...GluIleLysThr 206  
3336 GAGAAATGGAAACCTTAAGGCACTATGAGTGTGTGAAAGTAAAGAT 3385  
207 PheArgLeuLysLeuGlnAsnLeuGlnThrValLysAspGlnAla... 221  
3386 GCTGACACAAACATGAACCTTAAGAAACTATTAAGGATAAGAGCAACT 3435  
221 ..... 221  
3436 GCTTCATGAGAAAAATACTTTTTCAGCAATGACACTATATTCGGA 3485  
221 ..... 221  
3486 TTACACCTTTTACAGACTCGCTCTCCCTCAAAATTAAGTTGAAGGAAAC 3535  
222 .....HisLysLeu... 224  
3536 TCTCAAGCCCCATAGAATCAATGACTACACAAATTAATATAGCCCTTGC 3585  
225 .....ArgGlnAsnIle... 228  
3586 TACAGAAAGAAACACATTATGCTGTCTAGACACTAAAGAAACAGTC 3635  
229 .....Ala 229  
3636 TCAAGGACCAAGTTATGATTGAAACACTCAACTCAAAAGCTTCAACGA 3685  
230 GlnAspGlnLysSerAspAlaSerLysSer... 240  
3686 CAAGTATTGAAGTCTGATCTCCAGAAAGCCCAAGCAAGACTTGAAGA 3735  
241 .....GlnMetGlnLysGluLysGluLysIleC 250  
3736 AGGAGAGCTTAACTGCTTTTGGAGATGAACTACTGAAGGACACCTTA 3785  
250 ySgLTyThrgLysGluLysIle...LeuGlnMetGluThr...Ser 262  
3786 CTGACTACAGCTGCTATAGAAATGAGCACTGAGAACTGTGAAGT 3835  
263 LeuAspGluLeuArgArgLeuGlnLysGlnIleAspIleLysAlaThrG 279  
3836 ACAGAAAAACTCCAAACACTTCACAGAGAGATGAATAATATTACTATGA 3885  
279 uArgSerThrLeuLeuThrGln...GlnHisGluL 290  
3886 AAGGATGAGCTTCAAGCAACCTTGAAGACCTGAAGACAGACATGATA 3935  
290 ySLeuAlaIle...LeuSerGlnGlnAsnGlu...AspThrAsp 302  
3936 GCCTAAACAAAGACCTTATGTAACCACTTGAAGCACTGAATTAAGAAACAA 3965  
303 GlnGluLeuMetGluTrpGlnThrLysPheGluGluArgIleAlaLeu 319  
3986 GATGAATTAAGGCGCTGCCAGAAAGACTTAAGACACAGAAAGCAACTGCT 4035  
319 uGluThrLysLysSerLysLeuValArgAspMetAspAspGluAspSer 336  
4036 TGAATGCTTGAACACAGCTT...TTGATTTGTCTGTAGAA 4076  
336 yIserSer...ValLeuSerLysGlnAsnSerGluLeuThr 348  
4077 TTTTCATCCAAACCAACATGATGCACTGCTTAACCAAGAAAGTGTCAATG 4126  
349 HisGluIleGlyLysLeuGlnAlaGlu... 357  
4127 GGTAAATTAATTTGTTTCAAAAGTAAATGCTGCGTGTGAAGATGA 4176  
358 .....AlaAspAlaHisLeuThrMetL 365  
4177 GCTCAAAACATCTTGAAGCATTAAGTTCAGAACTGGAGCTACTGCTG 4226  
365 yHisGluArgAspSerAspIleLysAsnIle... 375  
4227 CTCAATGAAATCTGTGAGGAGGAGAAACCTTGAATCACAACAAACCTC 4276  
376 .....CysThrLysHisAsnLeuGlyProValProGluHisPropeth 390  
4277 AATGGCTTGAAGAGATATTTGGCAATCTGAAGAAACCAAGTGT 4326  
390 rAsnAspValAlaMetAsnLeu...ThrAsnArgIleLysAla 404  
4327 GAAGTCCATGTTGGAGAACTTAAGAGAACACAAATTAAGCTCAAGAAC 4376  
404 rG...LeuSerSerLeuGlnAsnAsp... 411  
4377 AAGCAGAGCAATTTCTAAGAAATTAATCAATTCAGTTTGAAAGAGCTG 4426  
412 .....LeuLeuAsp... 415  
4427 TTCAGTGGTTCACAGAACTGCTAGATGATAGAGTCCCTGAAGACCA 4476  
415 sLysLysSerAsnGlnAspGlnLeuAspValLeuTrpLysHisTyrLeuL 432  
4477 GCTAAAGCAGCAGAAAGAGCTGGAATTAAGGATAGAGATTTATTTG 4526  
432 ySile...AsnAlaArgTyrSerGluValAspGlyGlnIleGlnSer 446  
4527 AACTGTACAAACCTGCAACACCAATTAAGTGAAGGAAATTTGAAACT 4576  
447 LysIleGlnSer...MetSerGlyIleLeuArgArgLysAs 460  
4577 CCATTGCAAGCTGACCAATGAGAGAACAGCATTAATCGCGCTTCGAAGA 4626  
460 pLysGluLysGluArgAspAlaValGluLeuSerLysPheAsnL 477  
4627 AATGAGATTAAGTCTTGGGAAAACTTGAGCGGAATCAGTATTATC 4676  
477 euserArgIleAspGluArgGluArgHisMetGlnIleGluValGluArg 493  
4677 TGAAGAGATTCAGAAAGAAAGCTGGAACGTCTCAACAACTTGA... 4723  
494 LysThrLeuAlaLeuGlyGluArgAspTyrAspSerIleIleSerGlnLys 510  
4724 .....AspCysGlnGlu 4734  
510 sArgThrGluValTyr...SerLeuGlnGlnLysIleL 522  
4735 AGAGATGAGAGAGCTCGCTTCTATTAAGATGACTGCAACAGAGAGTGA 4784

```
522 yValleuLeuArgGluLysAspIleIleAsnArgAsnAlaAspGluArg 538
:: ||||| ||||| ||||| ||||| |||||
4785 AAGCTTGCTGCTGAGTGAAGAACTATTATTCTAAAGAGATATTGACACAGC 4834
539 ValLys..... 540
4835 CTAAAGCATCATCTCAGATCTCAGCTCAGTGCAGAAAACACAGCAGCA 4884
541 LeuGluLeuLysLysAsp.....AlaL 548
||| ||| |||
4885 GCTACAGTTGGCTGAAGAACTCTGCAATCGCTCTCTGCAATTGTCCTCA 4934
548 euGluSerSerLysAspLysLeuAsnGluIleValAsnGluHisLysLys 564
:: ||||| ||||| ||||| ||||| |||||
4935 TAACCTCAAGAAAAGAACCTCTGCAGATTGTCATCTCTCTGAGAGAA 4984
565 LysIle.....LysLysValLeuArgGlyArgAsnProheG 577
||| ||| ||| ||| ||| |||
4985 AAGATATTGTTATTAACTGAGAAATTCATCAAAAACATAATGACAGCA 5034
577 uLysAspMet.....LysLysGluIleAsnGluAlaPheTTPProValA 592
||| ||| ||| ||| ||| |||
5035 AAAATTACTACATGAAAGAAATGAACCTTGAAACAGCTCAGTTGAGCTAA 5084
592 sPLySGluTyrAsnGluLeuArgSerLysSerGluGluAlaGluGlu 608
||| ||| ||| ||| ||| |||
5085 AGTCGAGGTGAGACATCTGATGATGATGATGATGATGATGATGATGAT 5134
609 LeuLysPheThrGlnSerLysValThrAspAlaArgGluGluLeuThrLys 625
||| ||| ||| ||| ||| |||
5135 CTGAGTCTCTGACGATGAGAAACATGATCTGAAACAACTCTCTGCG 5184
625 sLeuArgArgAspMetAspAlaLysArgArgPheLeuAspSerLysLeuG 642
||| ||| ||| ||| ||| |||
5185 TCTTAACAGCAGATGCAAGTACTGCTCAAGAAAAGAAAGCTGCAAC 5234
642 InSeTlleuGlnIleSerAlaAsnValAspMetPheProLysValLeu 658
||| ||| ||| ||| ||| |||
5235 AAACCCATGAACTTAACTGAGTGCAGCTGAGTGCAGCTGAGTGCAGT 5284
659 GlnAspAlaMetAsnLysArgAspGlu..... 667
||| ||| ||| ||| ||| |||
5285 GAATGGGTTGATTTTAAATGAGCGCAGCAAAAGCCACTAAAGA 5334
667 ..... 667
5335 GCATGTCTGTAATGAGATTAAGAACTTGAGCAAGCAGCAGCAGCAG 5384
668 ..GlnLysArgLeuGluAsnPheAlaAsnGlyMetArgGluMetLeuAla 683
||| ||| ||| ||| ||| |||
5385 TTCAATGTGAGATAGAGAGCTTATGCAAAAGCTTAAAGATTAAGAGCTCA 5434
684 ProPheGluHisLeuAlaArgLysAsnHisValCysProCysCysGluArg 700
||| ||| ||| ||| ||| |||
5435 GCGCTGGAACTTTAAAGAACTGAA.....CAAA 5466
700 gAlaPheThrProAspGluGluAspGluPheVal.....LysLysGlnA 715
||| ||| ||| ||| ||| |||
5467 AGTAATTATTCGAAACCAAGAAATGGAATGTTATGCTGAAATGAGAG 5516
715 rMetGlnAsnSerSer.....ThrAlaGluArg..... 724
||| ||| ||| ||| ||| |||
5517 AATTGAAAAATAGCCAGAGACTGTAAATGCTGAGAGGAGCCAGCTGCAA 5566
725 .....SerLysAlaLeuAlaMetGluSerSerAsn..... 724
||| ||| ||| ||| ||| |||
5567 GACGACCTTAGGAGAGTGTGAAATGCTCATTTGAAATCAGATGATCT 5616
735 .....AlaGluAlaLeuPheGlnGluLeuAspLysLysArgThrT 748
||| ||| ||| ||| ||| |||
5617 AAGAAAGCTCAAGAGCATGTGACAGCAGCAAGAAATTAAGTTACAGAGAC 5666
748 leryThrAspAlaTyrValLysLeuValGluGluThrIleProLeuAlaGlu 764
```

```

6564 AACGATGATGATTCAGAGATTCAAG.....GAATCTTCAGAGAA 6607
990 LeuGlnGlyArgHisVal.....LeuCysHisSerGlnLeuGlnSe 1003
6608 ...CAGCTCCCATCTAAGTGAAGATTATGATATGAAAGCTGAAC 6654
1003 rCysMetAlaLysGln.....ArgHisSerAlaG 1014
6655 TCAATGCTCAGCAACAGAAAGACATCAACCAACCTGCAGAGA 6704
1014 lueuAnLysSerLysGlnLeuGlnGlnGlnGlnLysArg 1030
6705 AAGTTAAGGAAGTGATGACATATTCACACTTATCTCTAAAGAA 6754
1031 AsnLeuAsp.....AsnLeuLysTyrArgLys.. 1040
6755 CAGCTGACCAAAATACAGATGAGACTAAGAAATGAAGCTCAGAACTA 6804
1041 .....ThylsAlaAspVal..... 1045
6805 TGAACCTGCGAAAGATGATATCATGAAAGAAATCTCAGTCTGC 6854
1045 ..... 1045
6855 GTTAAATGCAAGACGCTCAGCAGAGAAAGATGATGTTGCAAGAGT 6904
1046 .....GlnGlnLeuThrArgAspLleGlnSerLeuGlnGlnuar 1058
6905 ATGATATACCTGAGACAGAAACCAAGAAATACAGAGCTGATGAGAAA 6954
1058 GlnLeuSerLleGlySer.....LeuSerAlaLleGln 1070
6955 AATCTCGCTGTATTCAGACAGCAACACTTCTCAGCAGCTCTCTA 7004
1070 laaSpluLysArgHisSerGln..... 1078
7005 GTGAGCTTCAAGAAAGAACTGACACACAAACTTCATCTTAAATATA 7054
1079 LysGlnArgLeuAsnSerGlnPheAsnArgTPrGlnGlyThrLeuSer.. 1094
7055 AAGAAATCTGTCATCAGCTCTCAGATCTTGGCAGCTTGCAGAAC 7104
1094 ..... 1094
7105 TGACATGTTAAGCTAATACTCAACTGCAGACCTCTGAAACAAATTA 7154
1095 .....ValTyrGlnSerAsnLleSerLysHisLysGlnLeuLysLeu 1109
7155 AGTTGATACCGAAGCTGCTGCACTCAAGAAAGATCATAGCTTATCAA 7204
1110 SerGlnTyrLysAspLle.....GlnLysArgTyrThrAsnGlnPh 1123
7205 GATATATGGAAGACCTGCTGAGCAAAAGAGCATGATGAGCTGCG 7254
1123 eLeuGlnLeuLysThrGlnMetAlaAsnLysAspLeuAspArgTyr 1140
7255 ACTCCAACTGCAAGTGTGGAGCAGCAGCAAGAAATAGTGC..... 7297
1140 YrThrAlaLeuAspLysAlaLeuMetArgPheHisSerMetLysMetGlu 1156
7298 .....GATTCTCAGCTGAGAACTCAAGTTCTGTGAATTTGAA 7336
1157 GlnLeuAsnLysLle 1162
7337 TTTCTGATGATGATTACT 7354

```

seq\_name: /SIS1/gcgdata/geneseq/geneseq/NA2000.DAT.AA258980

seq\_documentation\_block:

ID AA258980 standard; cDNA; 7962 BP.

AC

AA258980;

XX

08-MAY-2000 (first entry)

DE Human cytoskeletal protein (HCYT) encoding cDNA (clone 1655208).  
 XX  
 DE Human cytoskeletal protein; HCYT; cell proliferation; immunological;  
 KW reproductive; nervous disorder; cytostatic; immunosuppressive; anti-HIV;  
 KW anti-diabetic; arteriosclerotic; dermatological; anti-inflammatory;  
 KW anti-infectivity; vasotropic; cardiant; ss.

OS Homo sapiens;

Key Location/Qualifiers  
 CDS 239..7567  
 /tag= a

WO200006730-A2.

10-FEB-2000.

30-JUL-1999; 99WO-US17167.

31-JUL-1998; 98US-0155185.

04-AUG-1998; 98US-0160081.

19-AUG-1998; 98US-0155228.

(INCY-) INCYTE PHARM INC.

Bandman O, Tang YT, Yue H, Corley NC, Guegler KJ, Azimzal Y,  
 Patterson C, Lal P, Baughn MR;

WPI: 2000-195297/17.

P-PSDB; AAY77575.

Human cytoskeletal proteins useful for diagnosing, treating preventing  
 cell proliferative, immunological, reproductive, developmental and  
 nervous disorders

Claim 9; Page 80-83; 84pp; English.

The invention provides human cytoskeletal proteins (HCYT) and nucleic  
 acids encoding the proteins. The HCYT polypeptides can be expressed by  
 standard recombinant methodology. The HCYT polypeptides, modulators and  
 antibodies are useful for treating or preventing a disorder associated  
 with decreased and increased expression or activity of HCYT in mammals.  
 The polypeptides are also useful for diagnosing HCYT activity disorders  
 such as cell proliferative, immunological, reproductive, developmental  
 and nervous disorders. Sequences AA258974-981 represent cDNA clones  
 encoding the HCYT polypeptides.

Sequence 7962 BP; 2240 A; 1699 C; 2600 G; 1223 T; 0 other;

alignment\_scores:

Quality: 449.00 Length: 1458  
 Ratio: 0.591 Gaps: 60  
 Percent Similarity: 52.126 Percent Identity: 20.645

alignment\_block:

US-09-538-396-2 x AA258980 ..  
 Align: seg 1/1 to: AA258980 from: 1 to: 7962

100 GlnLeuThrGlnLysAlaSerLysMetGlnPheLysAlaLleGlnSerVa 116

2372 CAGCTACATCAGAGAGCAAGCAAGCAGAGAA...GAAGCTTGGCCAGGCG 2418

116 lLeuGlnThrLleAsnProHisThrGlnLysValCysLeuSerTyrA 133

2419 AGTCCAGAGAGAGAGGCCCTTAGTACGAGAAAGCGGCTTAGAGGTGC 2468

133 rGcysAlaAspMetAspArgGlu.....IleProAlaLeuMet 145

2469 GGCTGACGCGCGTGAGAGGTGACGCGCAGAGACCTGCTGACAACTACAG 2518



```
4184 ATGGAACATATGAACTATGACATCC.....TT 4212
658 uGlnAspAlaMetAsnLysArgArgLysGlnLysArgLysGlnLysAsnPhe 675
1213 ACAGAGTCGGCTG...CGAGAGCAGAGCTACAGCGAATGGA.....G 4253
675 LAsnGlyMetArgGlnMetLeu.....Ala 683
4254 CCCAGGGTGAGCGAGAGTACTTCAGGCGACCAAGAGAACTGACAGACC 4303
684 ProPheGlnHisLeu..... 688
4304 CAGGTGAACAACCTGCAGAGCACTGTCTGTAAGAACCGAGGCTCAGGCAAG 4353
688 ..... 688
4354 TGGTGGTGGATCTGTGAGAGAACCTGAGAGCGGCTCGCTCAGCACTGA 4403
689 ..AlaArgLysAsnHisValLysProCysGlnArgAlaPheThrPro 704
4404 AGCTGAAAAATGAGAGAGTAGAGAGTGAGCGTGAAGAGCCAGGCTCTG 4453
705 ArgGlnLysArgLysPhe...ValLysGlnArgMetGlnAsnSerse 720
4454 CAAAGGCAAGGGCGAACTGAAGAGTGCGCCCAAGGAGGCTCTGCAAGAGA 4503
720 ThrAlaGlnArgSerLysAlaLeuAlaMetGlnSerSerAsnAlaGln 737
4504 TTGGCCCTCTGACCCAGACCCAGCTGAAGAGAGAGAGAGAGTGAGAGA 4553
737 LalaLeu.....PheGlnGlnLeuAspLysLeuArgThrLysAsp 750
4554 CTCTGCGGGGCAAAATCCAGAGACTGAGAGAGCAACGGAATGCGACAG 4603
751 AlaArgValLysLeuValGlnGlnThrLysProLeuAlaGlnLysAsnLe 767
4604 GCTGCTTGGAAATG.....CTGCTTGGAACTGAGAGAGAG 4661
767 uAsnGlnHisLeuAlaAspGlnSerGlnLysAlaGlnAlaPheAspArg 784
4662 GAACCAAGAGAGTAGATCTGCAGCAAGAAAGATTCAGAGCTAGAGAGT 4691
784 LeuLeuGlnValLeuAlaHisValGlnMet..... 793
4692 GTAGCTGTGTTAGACATCTGCCATGGCCCTCAGAGCGAGAGCAG 4741
794 .....AspArgAspAlaValGlnAlaLeuLeuGlnProThrAs 806
4742 AAGCTGACTGTGCAGAGGAGCAGATCAGAGCTCGAGAAAGATCGGGA 4791
806 PThrLAspArgHisValHisGlnLysGlnLysValLysGlnValG 823
4792 GACT...CAGAGAAACGCTTGGAGCATCAGCTCTA.....GAACCTG 4832
823 LAspArgLeuLysArgLysAspSerSerGlnArgGlnValLysSerLeu 839
4833 AGAAGAAAGACCAATGATGATCCAGAGAGGAGCAGGTTCAAGACCTG 4882
840 GlnLys.....LysGlnLysGlnLys 847
4883 AAAAAGAGTGGTGAATCTGGAATGCTGGCCCTGGAATCGAGAAA 4932
847 nPhe.....LeuGlnArg 852
4933 CCATCACAAGATGAGAGTCCAGAAAACATGACAGAGCTGAGAGGCC 4982
852 hrArgAspThrLeuLysValGlnValAsp.....AspLeuArg 864
4983 AGAGGAAACCCAGAGAGTGGCTTGAACCCACCTTACGCTGAGCTGAGA 5032
865 ArgGlnHisArgMetLeuAsn.....G 872
5033 GAAAGAGCCAGAGAGCTGCAGAGCAAAAGCAGCCAGATCCATGACTGGA 5082
872 uAspMetSerSerAlaGlnValArgThrPheHisAsnAlaArgGlnLysVal 889
5083 GAGCCACAGCACCGCTCTGGCAAGAGAGCTGCAGAGAGAGGAGCAGAGAG 5132
889 AlLysAlaSer...SerLysLeuArgPheGlnLysSerGlnGlnLys 904
5133 TGAAGTCTCAGCAGAGAACAGATGAGAGAGCTGCAGAGCGAGAAAGCAT 5182
905 LeuValLeuLeuAlaGlnGlnLysGlnLys...LeuLysValLysLys 920
5183 CTGACTCAGAGATCTCGAGAGAGAGACAGAGAGCTGATGCTGCAGAGGA 5232
920 sLeuLeuGlnLysLysLeuAspProLeuSerLysGlnLysGlnSerLeu 937
5233 GAGAGTTCAGAGTCTCGAGAGTACAGAGAGCCGCGAGACCAATGCTGG 5282
937 euGlnGlnLysAsnAlaLeuLysGlnLysLeuAspGlnGlnLysGln 953
5283 AGGAGAGCTGGAACAGATCAACCTGCTTGAAGAGAGCGAGCGCGGAG 5332
954 LeuAlaGlnArgLysArgGlnPheGlnGlnLysLeuAspAlaLeuGln 970
5333 CTGAGCC...ACTCAGAGGAGAGCTGATCAGAGAAAGGCGAGAGAGGAA 5379
970 gLeuAsnMetLysLeu...LysGlnLysLysAspSerLysLys... 983
5380 GGGCCCAAGTAAAGCACAGCGCGGAGCTTGAAGACATGACATGACTGATCC 5429
984 .....AsnGlnLysLeuLysGlnLysGlnLysArgHisVal... 995
5430 TGGCTGATTAAGAGAGAGAGAGTGAATGTCAGAGAGCATATCCATGAA 5479
996 .....LeuCysHisSerGlnLysGlnLysSerCysMetAl 1006
5480 CTCAGAGAGCTCAAAAGCACAGCTGAGACAGACAGCTCCAGGCTGCACAG 5529
1006 AlLysGlnGlnArgLysLeuAlaLeuAsnLysSerLysGln... 1020
5530 GAAGGTGAGTGAACACAGCTCTCTGCTCCAGCGAGACAGAGAAATAG 5579
1021 ..LeuLeuGlnGlnLysGlnLysLeuLysArgAsnLysAspAsnLeu 1036
5580 TGGTCTCTGCAGACAGCACTGCAGAGAGCAGAGAA...CAAGGAGACTG 5626
1037 LysArgArgLysThrLysAlaAspValGlnGlnLeuThrArgAspArg 1053
5627 AAGGAGCAGTCACTTCAAGTCAACTGATGAGGCCAGAGAGCCCTAGC 5676
1053 uSerLeuGlnGlnLysLeuSerLysLeuSerLysLeuAlaLeuLys 1070
5677 CCAAGAGGAGCAG...GAACTGAGAGCTCTGAGC 5708
1070 LAspLysLysArgHisSerGlnLysGlnLysGlnLysArgLysSerGlnPhe 1086
5709 AAGAACAGCAGAGAGCCAGAGAGAGAGAGAGGCTGAAGAAAGCA 5758
1087 AsnArgTrpGlnGlnLysLeuSerValLysGlnLysAsnLysSerLys 1103
5759 GAGCCCTCCAGAGAGCTGAG...CAAGCCATATGACACTGAA 5802
1103 sLysGlnGlnLysLysSerGlnLysArgLysArgLysArgLys 1120
5803 GAGAGCTGATGAGAGAGCTTCAAGAGCAAGAGAGAGAGAGAGAGAGTGG 5852
1120 hrAsnGlnPheLeu.....GlnLysLysThrGlnMetAla 1132
5853 AGGAGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGG 5902
1133 AsnLysAspLeuAspArgLysArgLysThrAlaLeuAspLysAlaLeuMet 1149
5903 CTGGGAGAGCTTAAGGCTGAGTCTCGGAGAACAGAGAAAGCTGTGGC 5952
```

```

1149 gphetisermetylmetgluile.....Asn 1160
1150 .....
5933 CCTCAGCAGCAGTGTGCTGACAGCAGCAGCAGCAGTGTGAGAGACCA 6002
1160 ysilellelysgluLeuTrpGlnGlnThrTy.....Arg 1171
1161 .....
6003 GGGCCCTCAGACAGCAGTGTGCTGAGCCAGCAGCAGTGTCTCAAGAACGG 6052
1172 GlyGlnAspIleAspTyrIleSerIleAsnSerAspSer..... 1184
1173 .....
6053 GACCAAGAGCTGGAAGCTCTGCGGAGCAAGTCACTCCCGGAGTCA 6102
1185 ..... GluGlyAlaGlyThrArgSerTyrSerTyrArgValMetGln 1200
1186 .....
6103 GGAGAGAGCTGCGCGCGCGGCTGAGGCTGTGAGAGAGGCGCTT... 6148
1200 hrgIAspAlaGluLeuGluMetArgGlyArgCysSerAlaGlnGln 1216
1201 .....
6149 ..GGCAAGGCTATGCTGCGCCGCAAGGGA.....GAGCAGCAT 6187
1217 ValLeuAlaSerIleuIleIleArgLeuAlaLeu..AlaGluThrPheCys 1232
1218 .....
6188 CTCTCGAGCAGGAGCAATTTGAGCCGAGCTGTGAGGAGCAGCAGTGCAC 6237
1232 sLeuAsnGlyGlyIleLeuAlaLeuAspGluProThrThrAsnLeuAspG 1249
1233 .....
6238 CCTGCAACCTCTCCCTGATGCTGCGCAGCAGCAGTGTGAGTGTGAG 6286
1249 LyrProAsnAlaGluSerLeuAlaAlaLeuLeuArgIleMetGlnAla 1265
1250 .....
6287 .....GAGCTCTGAGGATACAGAGAGT 6310
1266 Arg.....LysGlyGlnGluAsnPheGlnLeu.. 1275
1267 .....
6311 GAGATCCAGCAGCAGATCTCCGATACAGAGATGTGTCCAGCAGCTGCA 6360
1276 .....ValIleThrHisAspGluArgPheAlaHisLeuIleGlyGlnA 1290
1277 .....
6361 GCAGCAGCTTGTCCGAGAGGATGAGAGCTGAGACTGACGAGAGACGG 6410
1290 rGlnLeuAlaGluLeuTyrTyrArgValSerLysAspGlnGlnHis 1306
1291 .....
6411 AGCAGCTCTGAGAGATCTCTGCGCCAGCAGGCTCCAGAGATATG... 6457
1307 SerIleIleGluSerGlnGluIle 1314
1308 .....
6458 ...ATCCAAGAGAGCAGAGATCTG 6478
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:AA134291
seq_documentation_block:
ID AAT34291 standard; cDNA to mRNA; 6175 BP.
XX
AC AAT34291;
XX
DT 25-MAR-1997 (first entry)
XX
DE Coding sequence for smooth muscle myosin heavy chain SM1.
XX
KW Smooth muscle; myosin heavy chain; SM1 isoform; rabbit; arteriosclerosis;
KW gene therapy; mouse; SM2 isoform; retrovirus; adenovirus; testis;
KW associated adenovirus; coronary artery catheterisation; sclerotic artery;
KW
XX
OS Mus musculus.
XX
PH key Location/Qualifiers
FT CDS 105..6023
FT /tag= a
FT /product= SM1
XX
PN W09623069-A1.

```

```

XX
PD 01-AUG-1996.
XX
PF 25-JAN-1996; 96MO-JP00134.
XX
PR 25-JAN-1995; 95JP-0010085.
XX
PA (OSAP) OSAKA PREFECTURAL GOVERNMENT.
XX (VSS-) VESSEL RES LAB CO LTD.
XX Arakawa E, Hasegawa K, Ishiyama H, Matsuda Y, Oda S;
XX Sugawara M, Takahashi K;
XX MPI: 1996-362693/36.
XX P-PSDB: AAM00024.
XX
PT DNA encoding smooth muscle myosin heavy chain SM1 isoform - used in
PT a vector for gene therapy for reduction of re-stenosis following
PT coronary artery catheterisation
XX
PS Claim 5; Page 14-27; 42pp; Japanese.
XX
CC This sequence represents the coding sequence for the smooth muscle myosin
CC heavy chain SM1 isoform protein. This SM1 coding sequence was isolated
CC from a mouse embryo cDNA library using probes based on the rabbit smooth
CC muscle myosin heavy chain SM2 isoform. This sequence is included in the
CC gene therapy vector of the invention. The vector of the invention
CC consists of this sequence inserted into a retrovirus, adenovirus,
CC associated adenovirus or animal-expression plasmid vector (such as PCXN2
CC or PAGE208). The vector can be used in the gene therapy treatment of
CC arteriosclerosis, particularly for the reduction of restenosis occurring
CC following coronary artery catheterisation for widening of sclerotic
CC arteries.
XX
SQ Sequence 6175 BP; 1846 A; 1455 C; 1056 G; 1018 T; 0 other;
XX
alignment_scores:
Quality: 446.00 Length: 1494
Ratio: 0.599 Gaps: 71
Percent Similarity: 49.799 Percent Identity: 20.750
alignment_block:
US-09-538-396-2 x AAT34291 ..
Align seg 1/1 to: AAT34291 from: 1 to: 6175
13 ArgSerPheAspProAspAsnLysAsnValIleThr..... 24
1884 AAGAACATGACCCCGCTAAATGACATGTGACATCTCTCAATGCCCTC 1933
25 .....PhePheLysProLeuThrLeuIleValG 34
1934 CTTGACAGATTGTGGCTGACCTGTGAGAGATGTGAGACCCGATTTGG 1983
34 LyrProAsnGlyAlaGlyLysThrThrIleIleGluCysLeuLysLeuSer 50
1984 GGCTGAGCCAGATGGCCAGATGACT..... 2009
51 CysThrGlyGluLeuProProAsnSerArgSerGlyHisThrPheValHI 67
2010 .....GAGAGTCACTGCGCAGTGTCTAAAGACC..... 2039
67 sAspProLysValAlaGlyGluThrGluThrLysGlyGlnIleLysLeuA 84
2040 .....AAAAAGGCATGTTCCGACCGCTGGAGACAGCTC..... 2072
84 rPheLysThrAlaAlaGlyLysAspValValCysIleArgSerPheGln 100
2073 ..TACAAAGACGAGTTGGGAAACTGATGTCTACACTGTGCGCAATTCAC 2120
101 LeuThr.....GlnLysAlaSerLys 107

```

2121 GCTAACTTTCGCGCTGATCATCCCAACATGAGAGAGCTGCGCAA 2170  
107 smetglu...PheLysalaileglu.....SerValleug 118  
2171 GCTGATGATCTTCTAGTGTGTAACAGCTGCGCTGCACAGCGTGTGG 2220  
118 lmrthleasn.....ProHis..... 123  
2221 AAGCATCCGATCTGCGCTGAGGCTTCCCAACAGAGATTGCTTCCA 2270  
123 ..... 123  
2271 GAGTTCGGGCAACCTATGAGATCTGCGACGCAACCGCATCCCAAG 2320  
124 .....ThrgLysLysValCys...LeuserYtrgCysAlaasp 137  
2321 CTTCATGATGTAAGAACAGCCTCATCTTCATGATCAAGCCCTCGAAC 2370  
137 etaspatrgluileProalaLeu...MetGlyValSerLysAlaVal 151  
2371 TTGAC.....CCTAACCTGTACAGAGATTGGCAGAGCAAAATCTTC 2411  
152 LeugLuanValIlePheValHisGlnaspGluSerAntrProLeugl 168  
2412 TTCGGAAGGGGGTCTGCGCCCACTAGAGAGAGACA..... 2450  
168 nasProSerThreLysLysLysPheaspPlePheSerAlaThra 185  
2451 .....GACGTAAATAATCATGATCATGCTGCTTCC 2484  
185 rg.....TyrThrLys..... 188  
2485 AGCAATGTGTGCTGCTACCTGCGCAAGAGCCCTTCCACCAAGAGCGAG 2534  
189 .....AlaLeuLysValIleLys..... 194  
2535 CAACAGCTGACAGCATGAGGTATGTCAGAGAGAACTGCGCTCACT 2584  
195 LysLeuHisLysaspGlnMetGlnLysLysThrePheArgLeuLysL 211  
2585 TAAAGTCCGCAACTGGCAATGTGGCGCTCTCAC.....AAG 2625  
211 euGluasnLeuGlnThyValLysaspGlnAlaHisLysLeuArg..... 225  
2626 TAAAGCCATTGCTCCAGGTGACACGCGAGAGAGAGATGACGCCAAG 2675  
226 .....GluasnIlealagLaspGlnLysSeraspAlase 238  
2676 GAGAGAGATGACAAAGATCAAGAGCGCGACGCAAGAGCGAGACTGA 2725  
238 rLysSerGlnMetGlnLysLysLysLysLysLysLysLysLysLys 253  
2726 GTTGAAGAGGTGAGAGAGACACACTAGCTGCTGAGAGAGACTC 2775  
254 .....ArgGluLysLeuGlnMetGlnThreSerLeu...AspGluLeuArg 267  
2776 TGTGTCAGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2825  
268 ArgLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 284  
2826 GAGATGCGGGTCCGTTGGCAGCCAAAGAGAGAGAGAGAGAGAGATCT 2875  
284 utHrgLinsLysLysLysLysLysLysLysLysLysLysLysLysLys 299  
2876 A.....CATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2916  
300 .....Asp 300  
2917 GCCACCACTACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2966  
301 ThrAspGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 314  
2967 CTGGAGAGCAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3016  
314 uArgIleAlaLeuLeuGlnLysLysLysLysLysLysLysLysLysLys 329  
3017 AGAGAGAGTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3066  
330 .....MetaspaspGlnAlaSerYtrSerSerValLeuSerLysGlnasn 344  
3067 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3107  
345 SerGluLeuThrHisGlnLysLysLysLysLysLysLysLysLysLys 361  
3108 AAACCTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3145  
361 sleuThreLysHisGlnArgaspSeraspIleLysasnIleCysThrL 378  
3146 CTTAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3195  
378 yshLaspLysGlnProValProGlnHisProPheThrAspValAla 394  
3196 AGCAT.....GAGTCTATGATCTCAGAGCTGAG 3224  
395 MetasnLeuThrasnArgIleValAlaArgLeuser..... 406  
3225 GTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3274  
407 .....SerLeuLuanaspLeuLeuaspLysLysSerasnGluasp 421  
3275 GAGGAACTGGAGGGGTGATGCTCAGTACTTCATGAGAGAGAGTGT 3324  
422 .....GlnLeuaspVal.....LeuTrpLysHisThrLeu 431  
3325 TGCAAGCCAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3374  
432 LysIleasnAlaArgYtrSerGluValaspGlnIleGln..... 445  
3375 GAGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3424  
446 .....SerLysIleGlnSerMetSerGlyLeu..... 455  
3425 CAATGCCCTAAAGAGATTCGAGAGAGAGAGAGAGAGAGAGAG 3474  
456 .....ArgArg 457  
3475 AAGAGAGACTGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3524  
458 ArgLysaspLysGlnLysGlnArgaspAlaIleGlnLysLeuSerLys 474  
3525 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3574  
474 sPheasnLeuserArgIleaspGluArgGlu.....ArgHisM 487  
3575 TACGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3624  
487 etGlnIleGlnValGlnArgLysThrLeuAlaLeuGlnLysArgaspYtr 503  
3625 AGGAGCTACAGCTGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 3674  
504 aspSerIleLeuser.....GlnLysArgThrGlu...ValTyrse 516  
3675 GAGGCCAGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3724  
516 rLeuLinsLysLysLysLysLysLysLysLysLysLysLysLysLys 533  
3725 ACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3774  
533 rg.....AsnAlaasp.....GluArg 538  
3775 AAAGCAAGCAGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3824  
539 ValLysLysGlnLysLysLysLysLysLysLysLysLysLysLysLys 555  
3825 CCGTCTCTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3874

```

555 uasnluilevalasnluhlslyasplysilelysvalleuarg 572
      |||
      |||
      |||
3875 GGAGGTGACAGTGCAGATCTGCTCCAGTCCAGTGCAGTGGAGCGCTG 3924
      |||
      |||
      |||
572 lyaasapnropheglulysasphetylsgluileasnluhlslyas 588
      |||
      |||
      |||
3925 CCCGGGTGACGTGACGACAGGTCCACAGCTCAGCAATGAATGAGT 3971
      |||
      |||
      |||
589 trprovalasplysgluityrasnluileuargserlyserserl 605
      |||
      |||
      |||
3972 GAGAGTGTACATGCTGCTCAATGAGGACAGGCAAGCCATCAACT 4021
      |||
      |||
      |||
605 agluigluileuysphethrlnserlysvalthralspalaargluc 622
      |||
      |||
      |||
4022 GCGCAAGATGTGCTTCCCTTGGATCCACCTTCAGGACACCCAGAGC 4071
      |||
      |||
      |||
622 lnleu.....thrlyseuarg 627
      |||
      |||
      |||
4072 TGCTCCAGAAAGAAACCCGCGAAGCTCAATGTCTCAACAGCTGCT 4121
      |||
      |||
      |||
628 Argasphetaspalalysargargpheleuaspserylseuargln 644
      |||
      |||
      |||
4122 CAG...TTGGAAGATGAAGACAGACCTGCAGCAGCAGCTGATGAGA 4168
      |||
      |||
      |||
644 eLeuGlIleserialasnyalaspmetpheprolyvalleuaglnas 661
      |||
      |||
      |||
4169 GATGGAGGCTAAGCAAAACCTGAGCGCCATGTCTCAACCTGAACAT 4218
      |||
      |||
      |||
661 lametasnlyasargaspoluglnlyasargluegluasnphealas 676
      |||
      |||
      |||
4219 AGCTCTCA...GACTTAAGAAAGAGCTGCAAGACTTGTCAAGTACC 4262
      |||
      |||
      |||
677 .....Glymetarglueuileuapropheglu 687
      |||
      |||
      |||
4263 ATCAGAGCATGAGAGGAGGAGAGAGGTTACAGAAAGATGAGAGG 4312
      |||
      |||
      |||
687 sLeuAlaarglysasnhsvalcysprocyasgluarglaphethr 704
      |||
      |||
      |||
4313 CCTCAGCCAGCAGTATGAG.....GAGAAGCGCGCTGCT 4347
      |||
      |||
      |||
704 roaspogluliaspdluphevalyslysglnargmetglnasnsers 720
      |||
      |||
      |||
4348 ATGCAAACTGGAGAA...ACCAAGAACAGCTCCAG..... 4382
      |||
      |||
      |||
721 ThrAlaIuargserlysAlaLeuAlametgluserSerAsnAla 737
      |||
      |||
      |||
4383 .....CAGGAGCTGATGACCTGCTGCTGCTGCTGCTGCTGCTGCT 4426
      |||
      |||
      |||
737 aleuphegluInleuaspIysleuargThrIleTyraspAlaTy 754
      |||
      |||
      |||
4427 ACTGTATCCAACTCGAAAAGACAGACAAATTTCAC.....C 4467
      |||
      |||
      |||
754 yslleuValgluIuThrIleProleuAlaIuIysAsnleuasnln 770
      |||
      |||
      |||
4468 AGTGTAGCTGAG.....GAGAAGAACATCTCTCCAG 4502
      |||
      |||
      |||
771 leuAlaspIusergluIysAlaIu..... 779
      |||
      |||
      |||
4503 TATCGGATGAGAGACCGAGCTGAGACAGACCGCAAGGAAAGAGAC 4552
      |||
      |||
      |||
780 .....AlaPheaspIleuargIuValleu 789
      |||
      |||
      |||
4553 AAAGCTTTGTCTAGCCCGGCTGAGAGAGCCCTGGAAGCCAAAG 4602
      |||
      |||
      |||
789 lahIstvalglu.....MetaspargaspAlaValgluAla 801
      |||
      |||
      |||
4603 AAGGAGCTGAGAGACCAACAGATGTCAAAAGCTGAGATGAAACCT 4652
      |||
      |||
      |||
802 leuGlInProthrIasprThrIleAsparIghIstvalhIsgluile 817
      |||
      |||
      |||
4653 GTCAGCTCCAGAGATGATGTAGCAGAGACGTGCTGAACCTGGAAT 4702
      |||
      |||
      |||
818 .....leuValysgluValgluasplleuIuTyralleuAsps 831
      |||
      |||
      |||

```

```

4703 CAAGGCTCCTTGGAGACCCAGATGGAAGATGAATAAAACCAAGCT 4752
      |||
      |||
      |||
831 erSergIlyargIlyValIysSerleuIuIu.....IleGlInleu 845
      |||
      |||
      |||
4753 AGTGAGAGATGAGCTGAGGAGCACTGAGATGCCAAAGCTCGGCTAG 4802
      |||
      |||
      |||
846 leuasnpe..... 848
      |||
      |||
      |||
4803 GTCAACATGACAGCCCTCAGAGCCAGTTTGAAGCCGATCTCAGGCT 4852
      |||
      |||
      |||
849 .....leuGlInargThrIaspr 855
      |||
      |||
      |||
4853 GGATGACAGATGAGAGAAAGAGAGGACACTTACAGCGGACGTGCAC 4902
      |||
      |||
      |||
855 htleuilevalgluValaspaspleuargaspIlnhs..... 867
      |||
      |||
      |||
4903 AGTATGACAGCAATGTGAGATGAGAGAGAGAGAGCTGCGCGG 4952
      |||
      |||
      |||
868 .....ArgMetleuasnIuaspmetserSerAlaIuValarg 881
      |||
      |||
      |||
4953 GCAGCTAAGAAAGAGCTGGAAGGAGGAGGCTAAAGCTAGAGCTCA 5002
      |||
      |||
      |||
881 phIasnAla.....ArgIuIuIysVal.....LysA 891
      |||
      |||
      |||
5003 TGACTCAGCCATCAAAAGGAGGAGAGAGCCATCAAGCAGCTTGAA 5052
      |||
      |||
      |||
891 laseSerIleuargIuargphegluIys..... 900
      |||
      |||
      |||
5053 TGCAAGCTCAGATGAGAGCTTCCAAAGAGCTGATGATGCCCGTGC 5102
      |||
      |||
      |||
901 sergluIuIuIeValleuAlaIuIuIysgluIuIeVal 917
      |||
      |||
      |||
5103 TCAGAGATGAGATCTTGCACCTCAAAAGATGAGAG..... 5144
      |||
      |||
      |||
917 IgluIysleuIeugluIuIuserleuaspProleuSerIysglu 934
      |||
      |||
      |||
5145 AAAGCCAGAGGTGTGAGGAGAGACCTCATGCTC.....CAG 5184
      |||
      |||
      |||
934 luserleuIeuglnIuTyrasnAlaIeulysgluIysleuaspIu 950
      |||
      |||
      |||
5185 AGGACTGCGACAGCTGAGAGAGCTGCCAAGCAA...GCTGACTGAG 5231
      |||
      |||
      |||
951 TyrHisgluIeulagluargIysargIuphegluIuIe..... 965
      |||
      |||
      |||
5232 AAGGAGAGCTGCGGAG.....GAGCTGCTACACTGCTGACG 5272
      |||
      |||
      |||
966 .....AspAlaIeuglyargIleuasnMetIysIle 977
      |||
      |||
      |||
5273 AAGGATACACTGACAGATGAGAAAGCCGCTGAGGCAAGATC... 5318
      |||
      |||
      |||
977 lYTyrlleuaspserIysAsnIuIysleuIysgluIeuglnIarg 993
      |||
      |||
      |||
5319 .....GCCAACTAGAGAGAGCTGAGAGAGACAGGACGAC 5357
      |||
      |||
      |||
994 HisValleuCyHisSergIleuIeuglnserCyMetAlaIysgl 1010
      |||
      |||
      |||
5358 ATGAGGCGCAGATGATAGATAGCAAGGCAACACTGAGGCGTGA 5407
      |||
      |||
      |||
1010 gIleSeralagluIeuanIysSerIysgluIeuglnIyglnglu 1027
      |||
      |||
      |||
5408 ACTGAGCAATGCTGGCCACAGAACGACGACGCTAG..... 5447
      |||
      |||
      |||
1027 InleuIyargAsnIleaspaspaenleuIysTyragIyThrIys 1043
      |||
      |||
      |||
5448 .....AAGAAATGAGAGC 5459
      |||
      |||
      |||
1044 AspValgluIuIeThrArgaspIleIuserleuIuIuIargIe 1060
      |||
      |||
      |||
5460 GCAGCGCAACAGCTGAGCGGCAAGAAAGAGATGCCAACAAGTTG 5509
      |||
      |||
      |||
1060 uSerIle.....GlyserleuSerAlaI 1068
      |||
      |||
      |||

```



5510 GGAGTAGAAGTGTCTCAAGCAAGCTCAAGTCTCGCGCC 5559  
 1068 legluAlaAspLeuLysArgHisSerGlnGluLysGluArgLeuAnSer 1084  
 5560 TGGAGGCAAGATT.....GCACAGCTGGAGGAGCAGGTGAACAG 5600  
 1085 GluPheAsnArgTrpGlnGlyThrLeuSerValTyrGlnSerAnIle 1101  
 5601 GAGGCCAGAGAAACAGGGCCACCAGTCCGTGAAGCAAGAACAA 5650  
 1101 rLysHisLysGln.....GluLeuLysLeuSer. 1110  
 5651 GAGGCTAAAGAGAGTCTCTGTCAGAGTGAGATGAGCCAGATGGCAG 5700  
 1111 ..GlnTyrLysAspIleGluLysArgTyrThrAsnGlnPheLeuGlnLeu 1126  
 5701 AGCAGTACAAAGAGAGCAGCAGAGAAAGAAACCAAGGTCAAGCAGCTG 5750  
 1127 Lys.....ThrThrGluMetAlaAsnLysAs 1135  
 5751 AAGAGGCAAGTGTGAAGAGCAGAGAGAGTCCAGTGCATCAAGCCAA 5800  
 1135 PLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPheHis 1152  
 5801 CCGCAGGAAGCTGCAGCGGAGCTAGTAGAGCC.....A 5835  
 1152 erMetLysMetGluGluIleAsnLysIleIleLysGluLeuTrpGlnGln 1168  
 5836 CAGAGAGCAATGAGGCCATGGCCGTGAGTGAAGCCCAAGAGCAGCAA 5885  
 1169 ThrTyrArgLysGlnAspLysPheLysSerIleAsnSerAspSerGln 1185  
 5886 CTACAGAGAGAAACGAGGCTCATTTGTT.....CCTTCAG 5923  
 1185 uGlyAlaGlyThrArgSerTyrArgValAlaMetGlnThrGlyAla 1202  
 5924 AAGGCTGGGGGCGCT.....AGAGTTATGAAACACAGATG 5961  
 1202 sPaLaGluLeuGluMetArgLysArgCysSer 1212  
 5962 GTTCTGAAGAAAGAAATGGACGCTCGGAGCTCA 5993

seq\_name: /stds1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV21518

seq\_documentation\_block:  
 ID AAV21518 standard; cDNA; 5661 BP.

AAV21518;  
 17-AUG-1998 (first entry)

Rattus norvegicus mutant alpha-myosin heavy chain transgene.

Actin binding domain; alpha myosin heavy chain; bridging; mouse;  
 transgenic mammal; congestive heart failure; study; treatment; diet;  
 exercise; effects; identification; hypertrophic cardiomyopathy;  
 dilated or hypertrophic cardiomyopathy; acute aortic regurgitation;  
 tricuspid stenosis; constrictive pericarditis; hypertension;  
 acute infective endocarditis; ischemic heart disease;  
 primary myocardial disease; valvular disease; pericardial disease;  
 hyperthyroidism; anaemia; arteriovenous fistula; beriberi;  
 Paget's disease; transgene; ss.

Rattus norvegicus.  
 Synthetic.

Key Location/Qualifiers  
 CDS 1..5661

/tag-  
 /product- "mutant alpha-myosin heavy chain"

MO9813476-A1.

PD 02-APR-1998.  
 XX 26-SEP-1997; 97MO-US17296.  
 XX 26-SEP-1996; 96US-0026742.  
 XX (LEIN/) LEINWAND LA.  
 XX Vikstrom KL;  
 XX WPI: 1998-230690/20.  
 DR P-PSDB: AAM54241.  
 XX  
 PT Transgenic mouse models for congestive heart failure and  
 PT hypertrophic cardiomyopathy - used to study molecular and cellular  
 PT events; identify potential therapeutic agents; assess effects of  
 PT diet etc.  
 XX  
 PS Claim 39; Pages 43-53; 75pp; English.

The sequence is that of a mutant rat alpha-myosin heavy chain  
 gene which was used in the development of transgenic mammals,  
 specifically mice. They can be used as a model for studying congestive  
 heart failure (CHF) or hypertrophic cardiomyopathy. Such animals  
 are used to study molecular and cellular events associated with  
 CHF; to identify compounds for treating CHF, and in evaluating  
 effects of diet and exercise on CHF. Conditions associated with CHF  
 that can be evaluated this way are dilated or hypertrophic  
 cardiomyopathy; acute aortic regurgitation; tricuspid stenosis;  
 congestive pericarditis; acute infective endocarditis; ischemic  
 heart disease; hypertension; primary myocardial disease; valvular  
 or pericardial disease; hyperthyroidism; anaemia; arteriovenous  
 fistula; beriberi and Paget's disease.

Sequence 5661 BP; 1572 A; 1446 C; 1797 G; 846 T; 0 other;

alignment\_scores:

Quality: 445.00 Length: 1403  
 Ratio: 0.614 Gaps: 59  
 Percent Similarity: 51.675 Percent Identity: 19.743

alignment\_block:  
 US-09-538-396-2 x AAV21518

Align seg 1/1 to: AAV21518 from: 1 to: 5661

98 SerPheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGln 114  
 1774 TCCCTCCAGACAGTGTCTGCTCCACCGGAAATCTGAACAAGCTGAT 1823  
 114 uSerValLeuGlnThrIleAsnProHisThrGlyGluLysValCysLeu 131  
 1824 GACAACTGTAGGACCCATCCTCAC.....T 1852  
 131 erTyrArgCys...AlaAspMetAspArgGluIleProAlaLeuMetGly 146  
 1853 TTGTGCCCTGATCATCTCCCAATGACGGAAGCTCA..... 1890  
 147 ValSerLysAlaValLeuGluAsnValIlePheValHisGlnAspGluSe 163  
 1891 .....GGGCGATGAGCAACACCCCTGTCATGACACAGTCGATG 1931  
 163 rAsnTrpProLeuGlnAspProSerThrLeuLysLysLysPheAsp... 178  
 1932 CAACGAGTGTGTGAGGATATCCGATCTGTAGAGAGGCTTCCCAACC 1981  
 178 ..... 178  
 1982 GCATTCTTATGGGACTTCCGCGAGAGGTATGATTCGATCCGACCAACA 2031  
 178 ..... 178

2032 GCCATCCCTGAGGGCCCAATTCATTGATACGGGAAAGGGCTGAGAAGCT 2081  
179 ..... Aspie..... PheserAlatThra 185  
2082 GCTGGGCTCCCTGGACATTGACCAACAGTACAAAGTTGGCCACACCA 2131  
185 rg... TyrThrLysala..... LeuGlValIleLysLysLeuHISLys 198  
2132 AGGTGTTCTTCAGAGCGGGCTGCTGGCTGCTGGAGGATGCGAGAT 2181  
199 AspGlnMetGlnIleLysThr..... 206  
2182 GAGAGCGCTAGCGGCATCATCACCAATCCAGCTCAAGCCGAGGCCA 2231  
207 ... PheAlaGLeuLysLeuGlnLeuGlnThrValLysAspGlnAlaH 222  
2232 GCTCAATGCGCATTTAGTTCAAGAAAGTGTGGAGCCGAGCGCCCTGC 2281  
222 IsLysLeuArgGluAsnIleAlaGlnAspGlnIleLysSerAspAlaSer 238  
2282 TGGTTATCCAGTGGACATC.....CGCGCTTCATGGGG 2316  
239 LysSerGlnMetGlnIleLysLys.....G1 247  
2317 GTCAGAAATGGCCCGCATGGAAGCTCTACTTCAGATCAAGCCGCTGCT 2366  
247 ulysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuA 264  
2367 GAGAGCCGAGAGACAGAGAGAGATGGCCACACATGAAA.....G 2407  
264 spGluLeuArgArgLeuGlnGlyGlnIleAspIleLysAlaThrGluArg 280  
2408 AGGAGTTCGGGGAGTCAAGATGCCTAGAGAGTCTGAGCTCGCCGC 2457  
281 SerThrLeuLeuThrGlnGlnIleGlnIleLysLeuAlaLeuSerGluG1 297  
2458 AAGGAGCTG.....GAGGAAAGATGCTGCTCCCTGCTGCAGGA 2495  
297 uAsnGluAspThrAspGlnGluLeuMetGluThrGlnThrLysPhe... 312  
2496 GAAAGATGCTCAGCTCCAGTCCAGGCGGAAACAAAGACCTGGCAG 2545  
313 ...GluGluArgIle.....AlaLeuLeuGlu 320  
2546 ATGCCGAGAGGCGCTGCGACAGCTGATCAAGACAAAGATCCAGCTGGAG 2595  
321 ThrLysIleSerLysLeuValArgAspMetAspAspGluAlaSerTyrSe 337  
2596 GCCAGGTGAGAGAGATGACCGAGAGCTGAGAGCAGAGAGAGATGAA 2645  
337 rSerValLeuSerLysGlnAsnSerGluLeuThrHisGluIleGlyLysL 354  
2646 CCCCGAGCTCAGCGCCAAAGAGCGCAAGCTGGAAAGCAGAGTCTCAGAGC 2695  
354 euGlnAlaGluAla...AspAlaHisLeuThrMet..... 364  
2696 TCAAGAAATATTCGATGACTGAGCTGACCTGGCCAAAGTGGAGAAG 2745  
365 ...LysHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAs 380  
2746 GAAAAGCCAGCAGACAGAAACAAGTTAAAAACCTG...ACAAGAGAGAT 2792  
380 nLeuGluProValProGlnHisProPheThrAsnAspValAlaMetAsnL 397  
2793 GCCCGGG.....CTGAGCAGATCATTTGCCAAG 2821  
397 euThrAsnArgIleLysAlaArgLeuSerSerLeuGlnLysAsnAspLeuG 413  
2822 TCACCAAGAGAGAAAGCT...CTTCAAGAGGCCACACAGCAAGCCCTA 2868  
414 AspLysLysLysSerAsnGlnLysAspGlnLysAspValIleuThrLysHis 430  
2869 GATGACCTTCAGAGCTGAGGAGACAAAGTCAACACACTGACCAAGTCTAA 2918  
430 IleuLysIleAsnAlaArgTyrSerGluValAspGlnIleGlnInsert 447  
2919 AGTCAGCTGAGACAGCAGGTGATGATCTGGAGGATCCCTGGACAGG 2968  
447 ysIleGlnSerMetSerGlyIleLeuArgArgArgLysAspLysGluLys 463  
2969 AG.....AAGAGTGGCGCATGTGACCTGGAGCCA 2997  
464 GluArgAspAlaAlaGluValGluLeuSerLysPheAsnLeuSerArgI1 480  
2998 GCAAGCGGAAGCTGGAGGTGACCTGAGCTGAGCTGACCCAGAGAGACTCAT 3047  
480 eaSpGluArgGluArgHisMetGlnIleGluValGluAlaGlyThrLeuA 497  
3048 GGACTGGAGAAAGACAGAGCTTCAGCTGGAGGAAAGCTCAAG..... 3090  
497 lalauGluArgAspTyrAspSerIleIleSerGlnLysArgThrGlu 513  
3091 .....AAGAAAGATTGAC.....ATCAGTCAAGCAAGACATGAA 3126  
514 ValTyrSerLeuGlnGlnLysIleLysValLeuLeuArgGluLysAspI1 530  
3127 ATAGAGAGAC...GAGCAGGCCCTGGCCCTTCAGCTGCAGAAAGAACTGAA 3173  
530 eileAsnArgAsnAlaAspGluArgValLysLeuGlyLeuLysLysAspA 547  
3174 GGAAGAACCAAGCAGCATGCGAGAGCTGGAGGAGAGCTAGAGCCGAGAC 3223  
547 lalauGluSerSerLys...AspLysLeuAsnGluIleValAsnGlnHis 562  
3224 GCACAGCCCGGCCCAAGGTGGAGAAAGCTGCGCTCAACACTGAGCCGGAG 3273  
563 LysAspLysIleLysLysValLeuArg.....GlyArgAsnProh 576  
3274 CTGAGAGAGATCATGAGAGCTGAGAGAGAAAGCGGTGGGCCCACTGCTG 3323  
576 eGluLysAspMetLysGluIleAsnGlnAlaPheThrProValAspL 593  
3324 GCAGATAGAGATGATACAAAGAGCGGAGCGCGGATTCAGAAAGATCGGC 3373  
593 yagLutyrAnGlu.....Leu 598  
3374 GGGACCTGGAGAGACCCAGCTGCACATGAGGCCACAGCCGGCCCTG 3423  
599 ArgSerLysSerGlnLysAlaGlnGlnIleLysPheThrGlnSerLys 615  
3424 CGCAAGAGAGCAGCAGACAGCGTGGCCGAGCTGGGAGACAGATACAA 3473  
615 sValThrAspAlaArgGlnGlnIleuThrLysLeuArgArgAspMetAspA 632  
3474 TCTACAGCGGTGAGACAGAGCTGGAGAAAGAGAGAGCGATTCAAAC 3523  
632 lalysArgArgPheLeuAspSerLysLeuGlnSerIleLeuGlnIleSer 648  
3524 TGGAGCTGATGAGCTACCTCTCACATGAGAGATCTCAAGGCCAAG 3573  
649 AlaAsnValAspMetPheProLysValLeuGlnAspAlaMetAsnLys... 664  
3574 GCAAAGCTGGAGAAAGTGCCTCCGAGACACTGAGAGACAGCGAGATGAA 3623  
665 .....ArgAspGlnIleLysArgLeuGlnLysAsnPheAlaAsnG 677  
3624 CCGGCTGAAGTTGGAGAAAGCCAGCGCTCCCTCAATGACTTCACACAG 3673  
677 LysMetArgGlnMetLeuAlaProPheGlnHisLeuAlaArgLysAsnHis 693  
3674 AGCGAGCCCAAGCTGCAGACAGAGATGCGAGTGGCGGAGAACTGGAA 3723  
694 ValCysProCysCys.....GluArgAlaPheTh 703  
3724 GAAAAGAGGACACTGATTTGGCAGCTGAGCCCGGGGGAACCTCTCCATAC 3773

[illegible]

5340 GCGCATGGGAGTGGTGAAGGC.....ATGAGGAGAGCGAGCGC 5383  
 1266 rglYslyGlnGluAsnPhelGlnLeuIleValIleThrHisAspGlu... 1281  
 5384 GCATCAAGAGCTCACCTACAG.....ACAGGAGAACACAG 5421  
 1282 .....ArpHeAlaHisLeuIleGlyGlnArgGlnLeuIleVal 1294  
 5422 AAGAACTGTGTGGCTGCAGACGCTGTGACACAGCTGAGTTGAAGT 5471  
 1294 ULSTYrYrArgValSerLysAspGluAsn..... 1304  
 5472 GAAGGCTTACAGCGCCAGCTGAAGAGCGGAGAACAGCCACACCA 5521  
 1305 .....GlnHisSerIleIleGlyGlnArgGlnGlu 1313  
 5522 ACCTGTCCAGTTCGCCAGAGTGCACAGAGCTGATGAGCAGAGAG 5571  
 1314 IlePheAsp 1316  
 5572 AGCGCGGAC 5580

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AA086851

seq\_documentation\_block:  
 ID AA086851 standard; DNA; 8789 BP.

XX AA086851;

DT 27-FEB-1996 (first entry)

XX Human mitosis gene.

XX Cell cycle; M phase; mitosis; retinoblastoma; mitosis; cell growth;

XX Inhibition; ss.

XX Homo sapiens.

XX MO9511309-A2.

XX 27-APR-1995.

XX 24-OCT-1994; 94WO-US12162.

XX 22-OCT-1993; 93US-0141239.

XX (TEXA) UNIT TEXAS SYSTEM.

XX Lee W, Zhu X;

XX WPI: 1995-170229/22.

XX P-PSDB; AAR72826.

XX P Purified mammalian protein mitosis and agents that bind it and

XX PT Inhibit its action - used to promote cell growth or to inhibit cell

XX PS division and/or proliferation

XX Claim 8; Fig 8B; 61pp; English.

XX AA086851 is the human mitosis gene. Mitosis is involved in the

XX regulation of the mammalian mitotic cell cycle. Mitosis as with E2F-1

XX (see AAR72824) interacts with the retinoblastoma protein (the

XX retinoblastoma tumour suppressor gene product). Mitosis is first

XX synthesised at the G1/S boundary, it is then phosphorylated from S

XX through M phase, and during mitosis, is closely associated with the

XX centromeres/kinetochores at the mitotic spindle poles. Mitosis is

CC A further use is to control hyperproliferative cells, and so control  
 CC diseases such as psoriasis and breast cancer. It can also be used to  
 CC block gametogenesis of an immature gamete.  
 CC Sequence 8789 BP; 3169 A; 1517 C; 2069 G; 2034 T; 0 other;  
 SQ

alignment\_scores:  
 Quality: 439.50 Length: 1256  
 Ratio: 0.668 Gaps: 49  
 Percent Similarity: 52.389 Percent Identity: 19.904

alignment\_block:  
 US-09-538-396-2 x AA086851 ..

Align seg 1/1 to: AA086851 from: 1 to: 8789

51 CysThrGlyGluLeuProPheAsnSerArgSerGlyHisThrPheValHis 67  
 3808 TGCTTGGGGAGACAGTCCCAATACCAT.....TA 3839  
 67 sAspProLysValAlaGlyGluThrGluThrLysGlyGlnIleLysLeu 84  
 3840 TGAGCCT.....CCAGGGAGAGATTAACCCAGGCTCTTCAGAA.... 3879  
 84 rPheLysThrAlaAlaGlyLysAspValValCysIleArg....Ser 98  
 3880 .....TGCAATTCGTGAATTCGA 3897  
 99 PheGlnLeuThrGlnLysAlaSerLysMetGluPhe.....Lys 111  
 3898 TTTTCTGGCTCTTAATGCTTTGTAACCTATGAGATTTCTGGGGAATCAGA 3947  
 111 sAlaIleGluSerValLeuGlnThrIleAsnProHisThrGlyGluLys 128  
 3948 AGATATCCATTAATCTCACTGCGGGTAAAGACATCAATAGACAATTT 3997  
 128 aLcylsLeuSerTyrArgCysAlaAspMetAspArgGluIleProAlaLeu 144  
 3998 TGAGATTACTTCACTGATAGAGAGCCGTCACGAAAGTGAACGTTTG 4047  
 145 MetGlyVal.....SerLysAlaValLeuGlnAsnValIle 156  
 4048 CTAAATGAATGAAGAATTAAGCTCAAAATCATTACAGAGGTACA 4097  
 156 ePheValHisGlnAspLysSerAsnTrpProLeuGlnAsp...ProSerT 172  
 4098 ACTAATGACCAAAATGGAAGCATGATGAATGGAAAAATAGTTGGGG 4147  
 172 hLeuLysLysLysPheAspPheIlePheSerAlaThrArgTyrThrLys 188  
 4148 AACCTTAAGAAAGAAACTCAGATTTA.....ACTGAA 4179  
 189 AlaLeuGlnValIleLysLysLeuHisLysAspGlnMetGlnGluLys 205  
 4180 AAATGGAAATATTTCTTGTGATCCAGAGAGTTACTCAGAGATAGA 4229  
 205 sThrPheArgLeuLysLeuGlnAsnLeuGlnThrValLysAspGlnAlaHis 222  
 4230 AACTTCTGAAGGCTCAATCTGATTGAAGAATGCATGCAAGAAATCAAT 4279  
 222 IsLysLeuArgLysAlaGlnIleAlaGlnAspGlnLysSerAspAlaSer 238  
 4280 CA.....CCTGAAGATATGAGAAATGTCGCAAGGTGAATGACAG 4323  
 239 .....LysSerGlnMetGlnGlnIleuLysG 247  
 4324 TGAGAGAGAGATTTCTGTATGTGAAATGACCTGATGATGATCAGATC 4373  
 247 ULysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeu 264  
 4374 GGAGAAAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 4423

```

264 spgIleuArgArg..... 268
4424 AGTAGTTCACACAGACAGATGTTAGAAAAAGACAAATGAATAG 4473
269 .....LeuGlnGlyGlnIleAspIleuValThrGly 279
4474 CAGAGGTTATTGTCGCTTGAAGAAAGACTCAGTGCACAGAGTA 4523
279 uArgSerThrIleu.....LeuThrGlnGlnIleGlu 290
4524 GAGAAACCGCTTCGTGAGATTAGATCTATGTCAAAAAACCCACGG 4573
290 yIleuAlaIleuSerGlnGluAsnGluAspThrAspGlnIleuMet 306
4574 CACTGATCACTTGTCTCAAAAAATGAAGGAAAAACACAGACTTCAG 4623
307 GlnTPrpGlnThrLysPheGlnGluArgGlyIleAlaLeuLeuGlnThrLys 323
4624 TCTCATCAAAAGTGAAGTGTCTCCATTCATTCAGTGCAGCAGCAGAGT 4673
323 eSerLysLeuValArgAspMetAspAspGlnAlaSerTyrSerVal 340
4674 GAGGAAAAAGCGAACTCTTCAGACTTGTCTCTGATGTGAGTGAAGC 4723
340 euserLysGlnAsnSerGlnLeuThrHisGlnIleGlyLysLeuGlnAla 356
4724 TGTAAAAAGACAAACTATCTCCAGGAAAGCTGCAGAGTTTGAAAG 4773
357 GlnAlaAspAlaHisLeuThrMetLysHisGlnArgAspSerAspIle 373
4774 GACTCAGCAGCAGCTGTTGACAAATGTGAGCTGGAAGAACCAATGTC 4823
373 sasnIleCysThrLysHisAsnLeuGlyProValProGlnHisProPhe 390
4824 ACAACTGATTAAGAG..... 4839
390 hArgAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSer 406
4840 ..AAAGATTGCTTCTCAAGAAATCTGAAGCTTCAGGCCCTGAGCGCACT 4887
407 SerLeuGlnAsnAspLeuLeuAspLysLysSerAsnGlnAspGlnLe 423
4888 GAATCAGATTGTGAAGAACTGAATGCTCCAAAGCCTTGAGCGCGCACT 4937
423 uAspValLeuThrLysHisTyrLeuLysIleAsnAlaArgTyrSerGly 440
4938 GGTGAGAGAAAGGTGAGTCCGATTGAGCTGACACACAGCAGGAGAG 4987
440 aIAspGlyGlnIleGlnSerLysIleGlnSerMetSerGlyIleLeuArg 456
4988 TGCAT...CAGCTGAGAGAGGCAATCGAAGAACTGAGA.....GTTCCG 5028
457 ArgArgLysAspLysGlnLysGlnArgAspAlaAlaGlnValGluLeu 473
5029 ATTGAGGCCGATGAAGAGAGAGCTGCACATCGCAGAG..... 5067
473 rLysPheAsnLeuSerArgIleAspGlnArgGlnArgHisMetGlnIle 490
5068 .....AACTGAAAGACGCGAGCGG..... 5088
490 LuValGlnArgLysThrIleuAlaLeuGlnGluArgAspTyrAspSerIle 506
5089 .....GAGAAATGAT..... 5097
507 IleSerGlnLysArgThrGlnValTyrSerLeuGlnGlnLysIleLysVal 523
5098 .....TCACTTAAGATTAAGTGAGAA 5120
523 IleuLeuArgGlnLysAspIleIleAsnArgAsnAlaAspGlnVal 540
5121 CTTTGAAGGCAATGTTCAGATGTCAAGAAAGAAC...CAGGAGCTAGTGA 5167
540 yLeuGlnLysLeuLysAspAlaLeuGlnUserSerLysAspLysLeuAsn 556
5168 TTCTT.....GATGCC...GAGATTCCAAAGCAGAGTAGAG 5202
557 GlnIleValAsnGlnHisLysAspLysIleLysLysValLeuArgGlyArg 573
5203 ACTCTA.....AAACACAAATAGAGAGATGAGCCAGAGAGCT 5240
573 gAsnProPheGlnLysAspMetLysLysGlnIleAsnGlnAlaPheTrp 590
5241 GAAGTTTTCATTAAGACCTTGC..... 5265
590 roValAspLysGlnTyrAsnGlnLeuArgSerLysSerGlnAlaGln 606
5266 .....ACGTTAAGCTCTGAAAAAGAAACTGACAA 5295
607 GlnGlnLeuLysPheThrGlnSerLysValThrAspAlaArgGlnIle 623
5296 AAACAAATACAAAGAAAAACAAGT.....CAGTT 5324
623 uThrLysLeuArgArgAspMetAspAlaLysArgArgPheLeuAspSerL 640
5325 CTCAGAACTACAAAGTTACTCTCTCTTCAATTAAAGTCTGTTAAGAA 5374
640 yLeuGlnSerIleLeuGlnIleSerAlaAsnValAspMetPheProLys 656
5375 AGGAGCAAGCAGAGATACAGATC..... 5397
657 ValLeuGlnAspAlaMetAsnLysArgAspGlnGlnLysAlaGlnLys 673
5398 .....AAAGAAATCTAAACCTGACGTGAGAT 5426
673 nPheAlaAsnGlnLysMetArgGlnMetLeuAlaProPheGlnHisLeuAla 690
5427 GCTTCAGAAATCAGTTAAAGAGCTAAATGAGCAGTACAGACCTTTGCTG 5476
690 rGlyAsnHisValCysProCysGlyArgAlaPheThrProAspGln 706
5477 GTGACCAAGAAATTATGAAGCCACAGAACAGAGCTGAGCCACCAATA 5526
707 GlnAspGlnPheValLysLysGlnArgMetGlnAsnSerSerThrAlaGly 723
5527 GAGGAAAGG.....CATCAGCTGAGAAATACATTGAGAAAGCT 5564
723 uArgSerLysAlaLeuAlaMetGlnUserSerAsnAlaGlnAlaLeuPheG 740
5565 GAGACCCCGCTACAGCTGATGAAGAAAGACAGCTC...TGTGCTTAC 5611
740 InglnLeuAspLysLeuArg.....ThrIleTyrAspAlaTyrVal 753
5612 AACAACTGAGAGAAAGTAGCATCATGACAGATTACTTAAGGTAGAGTG 5661
754 LysLeuValGlnGlnThrIleProLeuAlaGlnLysAsnLeuAsnGlnH 770
5662 GAGAACTTGAAGAGAGCTAGATACCGCAGACCAAC...CAGAGACA 5708
770 sLeuAlaAspGlnUserGlnLysAlaGlnAlaPheAspLeuLeuGly 787
5709 TGCAGCTTTGAGCGAGAAATTCAAA..... 5736
787 AlIleuAlaHisValGlnMetAspArgAspAlaValGlnAlaLeuGln 803
5737 .....GAGAGAGTAGAGACCTAAAGCAAAATAGAGAGGTGACCA 5781
804 ProThrAspThrIleAspArgHisVal..... 812
5782 AGCTGAGAGCTGCGAATTGATGTTTACTTAAAGTCAAGAAAGA 5831
813 .....HisGlnIleGlnGlnLeuValLysGlnValGlnAspLeuG 826
5832 AAATCTGCAATGATGATTAACAAAGAGACCAAGCGAATATCTGATTAAG 5881
826 LuTyrAlaLeuAspSerSerLysArgGlyVal.....LysSerLeu 839

```



PS Claim 1; columns 23-40; 43pp: English.

CC This sequence encodes mitotin, a phosphoprotein necessary for the cell  
CC to enter mitosis. The protein's degradation is also necessary for the  
CC cell to advance into the next stages of mitosis. The mitotin protein,  
CC can be used to control the growth of cells. An anti-mitotin antibody,  
CC a mutant or a non-functional analogue of mitotin can inhibit the mitotic  
CC cell cycle by preventing the cells from entering the M phase, and over  
CC expression of mitotin or its functional equivalent, would inhibit the  
CC cycle by preventing cells from leaving the M phase. Antagonists to this  
CC protein can be used to control hyperproliferative cells in, (e.g.  
CC thyroid hyperplasia, Grave's disease, psoriasis, benign prostatic  
CC hyperplasia, Li-Fraumeni syndrome, breast cancer, sarcomas and other  
CC neoplasms, bladder cancer, colon cancer, lung cancer and various  
CC leukemias and lymphomas). Reintroduction or supplementation of lost  
CC mitotin function by introduction of the protein or nucleic acid encoding  
CC the protein into a cell can restore defective chromosome segregation,  
CC which is a marker of progressing malignancy. Malignant proliferation  
CC of cells can then be halted. The protein can also be used for the  
CC detection and diagnosis of hyperproliferative cells.

SO Sequence 8789 BP; 3169 A; 1517 C; 2069 G; 2034 T; 0 other;

Alignment scores:  
Quality: 439.50 Length: 1256  
Ratio: 0.668 Gaps: 49  
Percent Similarity: 52.389 Percent Identity: 19.904

Alignment block:  
US-09-538-396-2 x AAV09076 ..

Align seg 1/1 to: AAV09076 from: 1 to: 8789

51 CysThrGluLeuLeuProbaSerArgSerGlyHisThrPheValH1 67  
3808 TCCTCTGGGGAACAGTCCCAATACCAAT.....TA 3839  
67 sAspProLysValAlaGluThrGluThrLysGlyGlnIleLysLeuA 84  
3840 TGAGCCT.....CCAGGGAGATTAACCCAGGCTCTCAGAA.... 3879  
84 rPheLysThrAlaAlaGlyLysAspValAlaCysIleArg....Ser 98  
3880 .....TGCATTTCTGAATGTCA 3897  
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPhe.....Lys 111  
3898 TTTTCTGGCTTAAGCTTGTGTAATGATTCCTCTGGGAATCAGCA 3947  
111 sAlaIleGluSerValLeuGlnThrIleAsnProH1SerGlyLysVal 128  
3948 AGATATCCATATCTCACTGCGGGTAAGAAGACATCAATGAGAAATT 3997  
128 aLysLeuSerTyrArgCysAlaAspMetAspArgGluLeuProAlaLeu 144  
3998 TGAGATTACTCATGTGATGAGAGCCGTGACAGAAAGTTGAAGTTTG 4047  
145 MetGlyVal.....SerLysAlaValLeuGlnLysValIle 156  
4048 CTAAATGAAATGAAAGATTAGACTCAAAATCCATTACAGAGAGTACA 4097  
156 ePheValHisGlnAspLysSerAsnTyrProLeuGlnAsp...ProSerT 172  
4098 ACTAATGACCAAAATGAGAGCATAGATGAAATGAAATGAGTTGGGG 4147  
172 hLeuLysLysLysPheAspPheSerAlaThrArgTyrThrLys 188  
4148 AACCTTAGAAAGAAACTCAGATT.....AGTGA 4179  
189 AlaLeuGlnValIleLysLysLeuHisLysAspGlnMetGlnIleLys 205  
4180 AATTGGAAATTTTCTGTGTATCACAGAGTACTTCCAGAGATAGA 4229

205 sThrPheArgLeuLysLeuGlnAsnLeuGlnThrValLysAspAlaHis 222  
4220 AACTTTCGAAGCCCTCAATTTGTGATTGAAATGATCATGACATTAATCAT 4279  
222 sLysLeuArgGlnAsnIleAlaGlnAspGlnLysSerAspAlaSer 238  
4280 CA.....CGTGAAGATATGAGATATGTGCCCAAGTGAATGACAGC 4323  
239 .....LysSerIleMetGlnLysLeuGln 247  
4324 TGAAGAGAGATTCTTGATGTGGAATGAGCTGATGATGATGATC 4373  
247 uLysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuA 264  
4374 GGAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4423  
264 sPgluLeuArgArg..... 268  
4424 AGTGATTCAAACAGAGAGATGTTAGAAAAGACATGAAATGAG 4473  
269 .....LeuGlnGlyGlnIleAspIleLysAlaThrGln 279  
4474 CAGAGATTATTTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4523  
279 uArgSerThrLeu.....LeuThrGlnGlnHisGlnL 290  
4524 GAGAAACAGCTCTGCGAGAGATTAATGATCTATGCAAAAAAACACAGG 4573  
290 sLeuAlaAlaLeuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 306  
4574 CACTGATCAGTGTGTCTGAAAAATGAGAGAGAGAGAGAGAGAGAGAG 4623  
307 GluTrpGlnThrLysPheGluGluArgIleAlaLeuLeuGluThrLys 323  
4624 TCTCATCAAGAGTGTCTGATGATGATGATGATGATGATGATGATGAT 4673  
323 eSerLysLeuValAlaAspMetAspAspGlnLysSerIleVal 340  
4674 GAAG 4723  
340 eSerLysGlnAsnSerGlnLeuThrHisGlnIleGlyLysLeuGlnAla 356  
4724 TGTAAAGACAAATCTATCTCAGAGAGAGAGAGAGAGAGAGAGAGAG 4773  
357 GluAlaAspAlaHisLeuThrMetLysHisGlnIleArgAspSerAsp 373  
4774 GACTCAG 4823  
373 sAsnIleCysThrLysHisAsnLeuGlyProValProGlnHisProPhe 390  
4824 ACACTGAAATAAGAG..... 4839  
390 hPheAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSer 406  
4840 .AAAGATGCTGTGCAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAG 4887  
407 SerLeuGlnAsnAspLeuAspLysLysLysSerAsnGlnAspGln 423  
4888 GAATCAGATTATGAAAGAGTGAATGTCTCAAGGCTTGAGAGCCGACT 4937  
423 uAspValLeuThrPheLysHisTyrLeuLysIleAsnAlaArgTyrSerGlu 440  
4938 GGTGAG 4987  
440 aLysPheGlnIleGlnSerLysIleGlnSerMetSerGlyLeuArg 456  
4988 TGCAAT.....CAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5028  
457 ArgArgLysAspLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 473  
5029 ATTGAGCGGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5067

[illegible]

5709 TGCAGCTCTTGAGGCGCAGAAATCCAAA..... 5736  
5787 aLleuAlahIstValGInMeLstArgrAspAlaValGInAlLeuLeuGIn 803  
5737.....GGAAGAGTGAAGACCTTAAGAACCAAAATGAAAGGATGACCCA 5781  
804 ProthAspThrIleAspArgHisVal..... 812  
5782 AGCTGAGAGCTGTGGAATTAATGTGTGTACTATTAAGCTGACAAAAAGA 5831  
813.....HAsGInIleGInGInLeuValIAsGInValGInAspLeuG 826  
5832 AATTCGACAAATTAATTCACAAAAGACAAAGACGATATCTGAATTAG 5881  
826 IuTyTrAlaLeuAspSerSerGlyArgGlyVal.....LysSerLeu 839  
5882 AA..ATATAATATCATCATATTGAAAATATTTTGCAGAAAAAGAGCA 5928  
840 GInGInIleGInLeuGInLeuAsnPhLeuGInArgThrArgAspThrLe 856  
5929 GAAAAAGTCAAGTg.....AAGAAAAATCAAGAC 5960  
856 uIleValGInValAspAspLeuArgAspGInHisArgMetLeuAsnGInuA 873  
5961 TGGCATGAGAGATg....CTTCAAAACAACTTAAGAAGCTCATAGTA 6004  
873 sPheSerSerAlaGInValArgTrpHisAsnAlaArgGInGInu..Lys 888  
6005 GAGTGCACACCCCG.....CATATGACCAAGAAAGCCTGTAG 6042  
889 ValIAspAlaSerSerIleLeuGInArgPheGInLysSerGInGInGInuLe 905  
6043 GCCAAAGACAGAT..... 6057  
905 uValIleuAlaGInGInLysGInGInLeuIleValGInLysLysLeuT 922  
6058.....CTTAGAGTCAAGTAgAGTGTCTTAACCTTAGAGGCTCAgT 6100  
922 euGInGInSerLeuAspPro.....LeuSerLys 931  
6101 TGCTACAAAGCCTTGATGAGGCCAAAATTAATTAATGTGTTTGCAATCT 6150  
932 GInLysGInSerLeuGInGInLysAsnAlaLeuLysGInLysLeuAs 948  
6151 TCAGTAAATGGCCTCATTCACAAAGATAGAAAGTAGGCAAGACAGAACTGGA 6200  
948 pGInGInuTyHisGInLeuAlaGInuArgLysArgGInuPheGInGInuL 965  
6201 GAAAGAAAGATGAAGAAATAGTAGACCTGAAAATTAATTCAGAGCCAG 6250  
965 euAspAlaLeuGInArgLeuAsnMetLysIleLysGInTyTrLeuAspSer 981  
6251 AGCAGCTTGC..... 6261  
982 LysLysAsnGInLysLeuLysGInLeuGInGInLysArgHisValIleuCysHis 998  
6262.....TCTAAACTGTCCCAAGGTGGAAGAGACAGCAACCTTGGGA 6302  
998 sSerGInLeuGInSerCysMetAlaLysGInGInArgGInLeSerAlaGInL 1015  
6303 GGAGCAA.....AACTTGAACTGGAATGTGACAGTGGAT 6340  
1015 euAsnLysSerLysGInLeuLeuGInGInLysGInGInLysArgAsn 1031  
6341 TGAAGCAAGAAATCCAAAGTGTACAAATCCAA.....AAGGCTCT 6381  
1032 IleAspAspAsnLeuLys.....TyTrArgLysTrpLysAl 1043  
6382 TTGCAGAGACACATTAAGATGCTGCAGAGTCTTTCAGAAATCTAGAGAA 6431  
1043 aaSpValGInGInLeuThrArgAspIleuSerLeuGInuArgLeu 1060



[illegible]

PA (FOX-) FOX CHASE CANCER CENT  
PA (UYTE-) UNIV TECHNOLOGIES INT INC.  
XX Rattner JB, Yen TJ;  
XX PPI; 1996-287116/29.  
DR P-PSDB; AAR9795.  
XX  
XX  
PS Claim 6; Page 55-62; 72pp; English.  
CC  
CC A cDNA sequence (AAR34578) encodes a 372 kDa human kinetochore  
CC protein, CENP-F (AAR9795), that is detected by immunofluorescence  
CC microscopy only during the G2 and M phases of a cell cycle. The  
CC sequence was deduced from clones isolated. from a human breast  
CC carcinoma cDNA library in lambda gill by screening with serum contg.  
CC CENP-F antibodies and by 5'RACE. The cDNA can be used for prodn. of  
CC large quantities of recombinant CENP-F, or to design probes for the  
CC detection of CENP-F genes in test samples as a means of monitoring  
CC cellular proliferation.  
CC  
XX  
XX Sequence 10136 BP; 3767 A; 1768 C; 2396 G; 2205 T; 0 other;  
S0

Alignment scores:  
Quality: 439.50 Length: 1256  
Ratio: 0.668 Gaps: 49  
Percent Similarity: 52.389 Percent Identity: 19.904

Alignment block:  
US-09-538-396-2 x AAT34578 ..

Align: seg 1/1 to: AAT34578 from: 1 to: 10136

51 CysrhrglygluleuProProAsnserArgseGlyhsthrheValHI 67  
|||||  
5619 TGCTGTGGGGAACAGTCCCGATACCAAT.....TA 5650

67 sasProLysValAlaGlyIuThrgIuThrLysGlyIuLeuA 84  
|||||  
5651 TGACGCT.....CCAGGGGAAGATAAACCCAGGCGCTTCAGAA... 5690

84 rghelYsthrAlaAlaGlyLysAspValAlcysIleArg.....Ser 98  
|||||  
5691 .....TGCAATTCGTAATTCGA 5708

99 pheGlnleuthrGlnLysAlaSerLysMetGluPhe.....Ly 111  
|||||  
5709 TTTTGTGGTCTTAATGCTTTGTGACCTATGAGATTCTCTGGGAATCAGGA 5758

111 salatleGluSerValLeuGlnThrIleAsnProHsthrGlyIuLysV 128  
|||||  
5759 AGATATCCATATCTTCACTACGCGGCTAAAGAGACATCAATTGCAATT 5808

128 alCysLeuSerTyrArgCysAlaAspMetAspArgGluIleProAlaLeu 144  
|||||  
5809 TGAGATTACTCATGTATGAGAGACCGTCAGCAAAAGATTGAAGTTTG 5858

145 MetGlyVal.....SerLysAlaValLeuGlnAsnValII 156  
|||||  
5859 CTAAATGAATGAAGAATTAAGATCAAAATCCATTTTCAGAGAGTACA 5908

156 pheValIstGlnAspGluSerAsnTrpProLeuGlnAap...ProserT 172  
|||||  
5908 ACTAATGCCAAATTTGAAGCATGCATAGATTGGAAAAAATAGTTGGGG 5958

172 hrLeuLysLysLysPheAspAspIlePheSerAlaThrArgTyrThrLys 188  
|||||  
5959 AACTTAAGAAAGAAACTCAGATT.....AGTGA 5990

189 AlaLeuGluValIleLysLysLeuHstLysAspGlnMetGlnGluIleLys 205

```
||||| : : : : : ||| : : : : : ||| : : : : :
5991 AATTGGATATTTCTTGATCATCAGAGTTACTCCAGAGTAGA 6040
205 SThrPheanLeuLysLeuGluasnLeuGlnThrValLysAspGlnAlaH 222
||| : : : : : ||| : : : : :
6041 AACTCTGAGGCCCTCAATTCGATTAGAAATGCATGCAGATAAATCAT 6090
222 LsLysLeuArgLysAsnIleAlaGlnAspGlnGluLysSerAspAlaSer 238
||||| : : : : : ||| : : : : :
6091 CA.....CGTGAAGATATTGGAGATAAATGTGGCCAGAGTGAATGACACC 6134
239 .....LysSerGlnMetGluGlnLeuLysGln 247
6135 TCGAAGAGAGATTCTTGATGTGGAATAATGAGCTGAGTAGATCGATC 6184
247 uLysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuA 264
: : : : : ||| : : : : :
6185 CGAGAAAGCTAGCATGAGCATGAAAGCCCTTACCTGAGAGCTGACTAG 6234
264 sPGLuLeuArgArg..... 268
6235 AGTAGATTCAACAGAGAACTATTGTTAGAAAAGACAAATGAAATAAG 6284
269 .....LeuGlnGlyGlnIleAspIleLysAlaThrGln 279
6285 CAGAGAGTTATTGTCTGCTGCTTGAAGAAGACTCTAGTGTCTCACAAGTGA 6334
279 uArgSerThrLeu.....LeuThrGlnGlnHisGluL 290
||||| : : : : : ||| : : : : :
6335 GAGAAACACGCTTCGTGAGAAATGATGACTATGTCAAAAAACAACACGG 6384
290 yLysLeuAlaIleuSerGluGluAsnGluAspThrAspGluGluMet 306
||| : : : : : ||| : : : : :
6385 CACTGTGATCAGTGTCTGAAAAATGAAGAGAAAAACACAAAGACTGTGAG 6434
307 GluTrpGlnThrLysPheGluGluArgGluIleAlaLeuLeuGluThrLys 323
||||| : : : : : ||| : : : : :
6435 TCTCATCAAGAGAGTGTCTCATTCATTCAGTGTGCGCAGAGCAGAGCT 6484
323 eSerLysLeuValArgAspMetLysAspGluAlaSerLysSerValL 340
: : : : : ||| : : : : :
6485 GAAGAAAAAGACGGAACCTCTCAGACTTGTCTGCTGATGTGAGAGAC 6534
340 euSerLysGlnAsnSerGluLeuThrHisGluIleGlyLysLeuGlnAla 356
||| : : : : : ||| : : : : :
6535 TGTAAAGACAAAACATCATCTCCAGAAACCTCAGAGTTGGAAGAAG 6584
357 GluAlaAspAlaHisLeuThrMetLysHisGluArgAspSerAspIle 373
||||| : : : : : ||| : : : : :
6585 GACTACAGGCGCTCTTTGACAAAATGTGAGCTGGAACCAAAATTGC 6634
373 sAsnIleCysThrLysHisAsnLeuGlyProValProGluHisProPhe 390
: : : : : ||| : : : : :
6635 ACAACTGAATAAGAG..... 6650
390 hAsnAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSer 406
: : : : : ||| : : : : :
6651 ..AAAGATTGCTCTCAAGAAATCTGAAGCCCTCAGGCCAGACTGAGT 6698
407 SerLeuGluAsnAspLeuLeuAspLysLysSerAsnGluAspGlnLe 423
: : : : : ||| : : : : :
6699 GAATCAGATTATGAAAAGCTGAATCTCCAAAGCCCTTGAGGCCGCACT 6748
423 uAspValLeuThrLysHisLysLysLysIleAsnAlaArgLysSerGlu 440
| : : : : ||| : : : : :
6749 GGTGAGAAAGAGTGCATTCGATTCAGCTGAGCTCAACACAGAGAGAG 6798
440 aLAspGlyGlnIleGlnSerLysIleGluSerMetSerGlyLysLeuArg 456
||| : : : : : ||| : : : : :
6799 TGCAT...CAGCTGAGAGAGAGCATCGAAACTGAGA.....GTTCGC 6839
457 ArgArgLysAspLysGluLysGluArgAspAlaIleGluValGluLeu 473
: : : : : ||| : : : : :

6840 ATGAGCCGATGAAAAGACAGCTGCACATCGCAGAG..... 6878
473 rLysPheAsnLeuSerArgIleAspGluArgGluArgHisMetGlnIleG 490
: : : : : ||| : : : : :
6879 .....AACTGAAAAGACGCGCGG..... 6899
490 LuValGluArgLysThrLeuAlaLeuGlyGluArgAspTyrAspSerIle 506
||||| : : : : :
6900 .....GAGATGAT..... 6908
507 IleSerGlnLysArgThrGluValTyrSerLeuGluGlnLysIleLysVal 523
||||| : : : : :
6909 .....TCACCTTAAGATAAAGTTGAGAA 6931
523 lleuLeuArgGluLysAspIleIleAsnArgAsnAlaAspGluValL 540
||| : : : : : ||| : : : : :
6932 CCTTAAAGGAATTTCAGATGTCAGAGAAAAC...CAGGAGCTGATGA 6978
540 yLeuGlyLeuLysLysAspAlaLeuGluSerSerLysAspLysLeuAsn 556
||| : : : : : ||| : : : : :
6979 TTCTT.....GATGCC...GAGAAATTCAAACAGAGAGTAGAG 7013
557 GluIleValAsnGluHisLysAspLysIleLysLysValLeuArgGlyArg 573
: : : : : ||| : : : : :
7014 ACTCTA.....AAACACAAATGAAAGAGATGCCAGAAAGCT 7051
573 gAsnProPheGluLysAspMetLysLysGluIleAsnGlnAlaPheTyr 590
: : : : : ||| : : : : :
7052 GAAGATTTTGAAATTAGACCTGTCT..... 7076
590 roValAspLysGluTyrArgGluLeuArgSerLysSerGlnGluAlaGlu 606
||||| : : : : :
7077 .....ACGTTAAGGCTGAAAAAATAATCTGACA 7106
607 GlnGluLeuLysPheThrGlnSerLysValThrAspAlaArgGluIle 623
||||| : : : : :
7107 AAACAAATCAAGAAAAACAAGT.....CAGTT 7135
623 uThrLysLeuArgArgAspMetLysAlaLysArgArgPheLeuAspSerL 640
: : : : : ||| : : : : :
7136 GTCAGAACTAGACAAAGTACTCTCTCATTTAAAGTCTGTAGAAAGA 7185
640 yLeuGlnSerIleLeuGlnIleSerAlaAsnValAspMetPheProLys 656
||| : : : : : ||| : : : : :
7186 AGGAGCAGACAGATACAGATC..... 7208
657 ValLeuGlnAspAlaMetAsnLysArgAspGluGlnLysArgLeuGlnLys 673
||||| : : : : :
7209 .....AAAGCAAAATCTAAAACTGCAGTGAGAGT 7237
673 nPheAlaAsnGlyMetArgGluMetLeuAlaProPheGlnHisLeuAla 690
: : : : : ||| : : : : :
7238 GCTTCAGATCAGTTAAAGAGCTAAATGAGCAGTAGCAGCCTGTGTG 7287
690 rGlyLysHisValLysProCysCysGluArgAlaPheThrProAspGlu 706
: : : : : ||| : : : : :
7288 GTGACCAGAAATTAATGAGGCCACAGAACAGAGCTACACCAATA 7337
707 GluAspArgPheValLysLysGluArgMetGlnAsnSerSerThrAlaGln 723
||||| : : : : :
7338 GAGGAAGAG.....CATCAGCTGAGAAATAGCATTTGAAAGCT 7375
723 uArgSerLysAlaLeuAlaMetGluSerSerAsnAlaGluAlaLeuPheG 740
||||| : : : : :
7376 GAGAGCCCGCTAGAAAGTGAATGAGAAAGAGAGCTC...TGTGCTTAC 7422
740 InGlnLeuAspLysLeuArg.....ThrIleTyrAspAlaTyrVal 753
||||| : : : : :
7423 AACCACTGAGGAAGAGTGCATTCAGATTCATTCCTTAAGGTAGAGTG 7472
754 LysLeuValGluGluThrIleProLeuAlaGluLysAsnLeuAsnGlnHis 770
: : : : : ||| : : : : :
7473 GAGAACTTGAAAGAGAGCTAGAGATAGCCAGAGACAAAC...CAAGAGCA 7519
```

[illegible]

```

1043 aasvvalgluInleuthrArgspriIeGUsertleuGluInuArgLeuL 1060
      ::::::::::::::::::::|||
8243 TGACCTTGAATTGACAAAATGACAAAATGCTCTTTGTGAAAAA... 8288
1060 euserIleGlySerLeuSerAlaIleGluInaIleuAspLeuysArgHisSer 1076
      ::::::::::::::::::::|||
8289 .....GTAAACAAATACtCTGCAAGAAAGAACTAGCTGCAGAGGAAAG 8333
1077 gInGuL.....LysGluArgLeuAsnSerGI 1085
      ::::::::::::::::::::|||
8334 CATGAGATGGCACAGAAACAGCAGACCTGCAGAAAGAACTCAGTGCAG 8383
1085 uPheAsnArgTrpGlnGlyThrLeuSerValTYrGlnSerAsnIleSerL 1102
      ||||| |||
8384 GAAAAATGGCTAGCTGAGAGAGTGCAGTTACGTGGAGAAATATAACA 8433
1102 ySHsILysGInGluLeuLeuSerGlnTYrLysAspIleGluLysArg 1118
      |||||
8434 GCAGCAAAAGATCAATTGAG..... 8453
1119 TYrThrAsnGlnPheLeuGlnLeuLysTYrThrGluMetAlaAsnLysAs 1135
      ||| |||||
8454 .....GAGCTCACACrTGAATAATAGTAA..... 8477
1135 pLeuAspArgTYrTYrThrAlaLeuAspLysAlaLeuMetArgPheHis 1152
      ||| |||||
8478 .....TTGAAGAGAGCCCTTGATTCATGCATGCACAA 8505
1152 erMeTlyseMetGluGluIleAsnLysIleIleLysGluLeuTrpGlnGln 1168
      ||| |||||
8506 AAGACCAAGCTGCAAAAGAGGAAAGTGAAGAGAAATAGCTGAATAT 8555
1169 ThrTYrArgGlyGlnAspIleAsp.....TYrIleSerIleAsnSerAs 1183
      |||
8556 CAGCTACAGCGCTTCATGAAGCTGAAAGAAACACAGAGCTTGCCTTTGGA 8605
1183 pSerGluGlyAlaGlyThrArgSerTYrSerTYrArgValValMetGlnT 1200
      ::::::::::::::::::::|||
8606 CACA.....AACAAACAGATGAAGTAATCCACA 8637
1200 hrGlyAspAlaGluLeuGluMetArgGlyArgCysSerAlaGlyGlnLys 1216
      |||
8638 CATACCGAGAGAAATGACTCTCTAAAGAAAGATGCTCACTTACAGAAAG 8687
1217 ValLeuAlaSerLeuIle 1222
      ::::::::::::::::::::|||
8688 CTGAGATAGACCTTTTA 8705

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:AA158840
seq_documentation_block:
ID AA158840 standard; DNA: 580073 BP.
AC AA158840;
XX
XX 27-MAR-1997 (first entry)
XX
XX Mycoplasma genitalium genome.
XX
XX M. genitalium; DNA; DNA gyrase; origin of replication;
XX megabase shotgun sequencing method; open reading frame; ORF. ss.
XX
XX Mycoplasma genitalium.
XX
XX key Location/Qualifiers
XX CDS 852..914
      /tag=
      /label= M006
      /note= "Previously identified as MORF-20076, then
      encoded protein shows 27.59 percentage
      identity to thymidylate kinase (CDC8)
      from Saccharomyces cerevisiae"
XX
XX CDS 11252..12040

```

FT FT //tag- b  
FT FT /note- "Previously identified as MORF-20078, the  
FT encoded protein shows 35.43 percentage  
FT identity to the Bacillus subtilis hypothetical  
FT protein covered in accession number  
FT GB:D26185\_102"  
FT CDS 12069..12725  
FT //tag- c  
FT /label- MG010  
FT /note- "Previously identified as MORF-20079, the  
FT encoded protein shows 25.73 percentage  
FT identity to DNA primase (dnaE) from  
FT Clostridium acetobutylicum"  
FT CDS complement (13570..14247)  
FT //tag- d  
FT /label- MG012  
FT /note- "Previously identified as MORF-20080, the  
FT encoded protein shows 31.50 percentage  
FT identity to the ribosomal protein S6  
FT modification protein (rimK) from Escherichia  
FT coli"  
FT CDS complement (14396..15217)  
FT //tag- e  
FT /label- MG013  
FT /note- "Previously identified as MORF-19823, MORF-20080  
FT and MORF-20081, the encoded protein shows 33.04  
FT percentage identity to 5,10-methylene-tetra-  
FT hydrofolate dehydrogenase (fold) from E. coli"  
FT CDS 17474..19243  
FT //tag- f  
FT /label- MG015  
FT /note- "Previously identified as MORF-20084, the  
FT encoded protein shows 32.23 percentage  
FT identity to transport ATP-binding protein  
FT (msbA) from E. coli"  
FT CDS 26478..27344  
FT //tag- g  
FT /label- MG023  
FT /note- "Previously identified as MORF-20092, the  
FT encoded protein shows 45.96 percentage  
FT identity to fructose-bisphosphate aldolase  
FT (tssr) from B. subtilis"  
FT CDS 27345..28448  
FT //tag- h  
FT /label- MG024  
FT /note- "Previously identified as MORF-19826 and  
FT MORF-20093, the encoded protein shows 46.84  
FT percentage identity to GTP-binding protein  
FT from E. coli"  
FT CDS 36987..38978  
FT //tag- i  
FT /label- MG032  
FT /note- "Previously identified as MORF-20099, the  
FT encoded protein shows 26.82 percentage  
FT identity to ATP-dependent nuclease (addA)  
FT from B. subtilis"  
FT CDS 39242..39904  
FT //tag- j  
FT /label- MG033  
FT /note- "Previously identified as MORF-20100, the  
FT encoded protein shows 35.90 percentage  
FT identity to glycerol uptake facilitator  
FT (glpF) from B. subtilis"  
FT CDS complement (39873..40514)  
FT //tag- k  
FT /label- MG034  
FT /note- "Previously identified as MORF-20101, the  
FT encoded protein shows 48.13 percentage  
FT identity to thymidylate kinase (tdk)  
FT from B. subtilis"  
FT CDS 40543..41787  
FT //tag- l

FT FT /label- MG035  
FT FT /note- "Previously identified as MORF-20102, the  
FT encoded protein shows 30.71 percentage  
FT identity to histidyl-tRNA synthetase (hlsS)  
FT from Mycobacterium leprae"  
FT CDS complement (44751..46277)  
FT //tag- m  
FT /label- MG038  
FT /note- "Previously identified as MORF-20105, the  
FT encoded protein shows 46.83 percentage  
FT identity to glycerol kinase (glpK)  
FT from E. coli"  
FT CDS complement (46268..47422)  
FT //tag- n  
FT /label- MG039  
FT /note- "Previously identified as MORF-19831 and  
FT MORF-20106, the encoded protein shows 43.20  
FT percentage identity to glycerol-3-phosphate  
FT dehydrogenase (gnt2) from S. cerevisiae"  
FT CDS 49377..49643  
FT //tag- o  
FT /label- MG041  
FT /note- "The encoded protein shows 48.86 percentage  
FT identity to phosphohistidinoprotein-hexose  
FT phosphotransferase (phsH) from Mycoplasma  
FT capricolum"  
FT CDS 50060..51520  
FT //tag- p  
FT /label- MG042  
FT /note- "Previously identified as MORF-19832 and  
FT MORF-20108, the encoded protein shows 41.92  
FT percentage identity to spermidine/  
FT putrescine transport ATP-binding protein  
FT (putA) from E. coli"  
FT CDS 51525..52382  
FT //tag- q  
FT /label- MG043  
FT /note- "Previously identified as MORF-20110, the  
FT encoded protein shows 26.51 percentage  
FT identity to spermidine/putrescine transport  
FT system permease protein (putB) from E. coli"  
FT CDS 52366..53220  
FT //tag- r  
FT /label- MG044  
FT /note- "Previously identified as MORF-20111, the  
FT encoded protein shows 29.45 percentage  
FT identity to spermidine/putrescine transport  
FT system permease protein C (potC) from E. coli"  
FT CDS 54658..55605  
FT //tag- s  
FT /label- MG046  
FT /note- "Previously identified as MORF-20112, the  
FT encoded protein shows 36.60 percentage  
FT identity to sialoglycoproteinase (gcp)  
FT from Pasteurella haemolytica"  
FT CDS complement (56970..58310)  
FT //tag- t  
FT /label- MG048  
FT /note- "Previously identified as MORF-19834,  
FT MORF-20114 and MORF-20115, the encoded protein  
FT shows 43.02 percentage identity to signal  
FT recognition particle protein (ftn) from B.  
FT subtilis"  
FT CDS 58117..59079  
FT //tag- u  
FT /label- MG049  
FT /note- "Previously identified as MORF-20114 and  
FT MORF-20115, the encoded protein shows 44.78  
FT percentage identity to purine-nucleoside  
FT phosphorylase (deop) from E. coli"  
FT CDS 59083..59754  
FT //tag- v  
FT /label- MG050



495 ThrLeuAlaLeuGlyGluArgAspIleSerIleLeuSerGlnLysAr 511  
 260259 CACTTGAACACCAACCAAAATGAGTTT ..... 260285  
 511 GThrGluValIleYrSerLeuGlnLysIleLysValLeuLeuArgGlu 528  
 260286 ..... GAACAAA ..... 260294  
 528 YAspIleLeuAsnArgAsnAlaAspGluArgValLysLeuGlyLeuLys 544  
 260295 ..... CAGTCTGATAGCCTCTTAAGAGCTTGAACCTGAA 260327  
 545 LysAspAlaLeuGlnSerSerLysAspLysLeuAsnGluIleValAsnG 561  
 260328 TATTAAGGCTTACAA ..... CACAAGATTAATGAATTTAAAAATGA 260368  
 561 uHisLysAspLysIleLysLysValLeu...ArgGlyArgAsnProPhg 577  
 260369 AAGTCCCACTAAGACGAGAACTTTTAACCAAGAACGGAACTATTTG 260418  
 577 LuLysAspMetLysLysGluIleAsnGlnAlaPheTrpProValAspLys 593  
 260419 AAAA ..... CGCAGGAAATTCACACGCTTTAACCCACAGATCTTG 260462  
 594 GluTyrAsnGluLeuArgSerLysSerGlnGluAlaGlnGlnLys 610  
 260463 GAATATGACACCAACGT ..... GAGTCAGTCACTACTACTCA 260500  
 610 sPheTrpGlnSerLysValThrAspAlaArgGlnGlnLeuThrLysLeu 627  
 260501 AATAAGCAGCAATGAAGTAAACACACTCCAAACTTGAAGTATGCTA 260550  
 627 rArgAspMetAspAlaLysArgArgPheLeuAspSerLys ..... 640  
 260551 AAAAGAGCTAGACAAAGAAGAACTCTTAGTACACAAAAAAGTT 260600  
 641 ...LeuGlnSerIleLeuGlnIleSerAlaAsnValAspMetPheProLys 656  
 260601 GATAGTGAAGCAATCTTCACTCAAAAGAAAGTGTCAAGAACGTA 260650  
 656 sValLeuGlnAspAla...MetAsnLysArgAspGlnGlnLysArgLeu 672  
 260651 GGAGCTTGAAGACTATACCTGTGTAAAAAACAAGAACAGTCAAGG 260700  
 672 LuAsn ..... PheAlaAsnGlyMetArgGlnMetLeuAlaPro 684  
 260701 AAATGAACTGTTGTTTGTGAAACAGTTAAACACACCAAGCAGAT 260750  
 685 PheGluHis ..... LeuAlaArgLysAsnH 693  
 260751 TTTGAAGAAAGTGAAGCTAAACAACAGAGTTGTTGAACTAAACA 260800  
 693 sValCysProCysGlnArgAlaPhe ..... ThrProAspGlnG 707  
 260801 T ..... GCATGTGAACGTTCTTTATCAAACTTGAAGATTAAGAAA 260841  
 707 LuAspGluPheValLysLysGlnArgMetGlnAsnSerSerThr ..... 721  
 260842 AGATTTTAACCTAAAGCACACAGATGCGCATAGTTTCCCACTA 260891  
 722 ...AlaGluArgSerLysAlaLeuAlaMetGlnSerSerAsnAlaGluAl 737  
 260892 AAAACGTGAATGCAAG ..... AGTGCCTGATTTTGAACT 260926  
 737 AleuPheGlnGlnLeuAspLysLeuArgThrIleTyrAspAlaTyrVal 754  
 260927 AATGTGCAAAAT ..... 260939  
 754 YrSerValGluGluThrIleProLeuAlaGluLysAsnLeuAsnGlnHis 770  
 260940 ..... GAGTATGAAAACTTGCAACAGAA 260963

771 LeuAlaAspGluSerGlnLysAlaGlnAlaPheAspLeuLeuGlyVa 787  
 260964 AACCAAAATTTTCCACAGACGTAATCTTGAAGAAAGCGCGCT 261013  
 787 IleuAlaHis...ValGlnMetAspArgAspAlaValGluAlaLeuG 803  
 261014 TTTATCAACCGGTTACACAAAACGT ..... GAGGAGTTATTTAC 261054  
 803 LnPThrAspThrIleAspArg ..... 810  
 261055 AACCAAAAGAAACGCTGATCAGCTTACAAAAGCTTGAAGCAAGACG 261104  
 811 ...HisValHisGluIleGlnGlnValLysGluValGluAspLe 825  
 261105 TTTATCAACCAAAAGGACACAAAGGAGTTGTTCATCGTGAAGAA 261154  
 825 uGluTyrAlaLeu ..... AspSerSerGlyA 834  
 261155 AAGGAGATATTGGGCAAAAACCTCCAGATTTTCCCAACTCAGCTGA 261204  
 834 rGlyValLysSerLeuGlnGluIleGlnLeuGlnLeuAsnPhe ..... 848  
 261205 ACCCTCTAAAAATTTAGCTGAACGAGATGCAATCAAGTTTAAAGA 261254  
 849 ...LeuGlnArgPheArgAspThrLeuIleValGluValAsp ..... 861  
 261255 AAGGAGATGAAGCACTGAATAAACACACTGTTAATGATTTAATATGC 261304  
 862 ..... AspLeuArgAspGlnHisArgMetLeuAsnGluA 873  
 261305 TGAAGTTATCCAAAGCACTTGAAGTACCAACTCAACCAATCAACCA 261354  
 873 sPheSerSerAlaGln ..... ValArgTrpHisAsn 883  
 261355 AACGCACTGAATTCGCAAAACGCCAAACAAAGATTCGTATTTTCACAT 261404  
 884 AlaArgGlnGluLysValLysAlaSerSerIleLeuGlnLysPheGln 900  
 261405 GATTCCTAAAAAACTCAATGAGTGAACCTT ..... AGCTTCAAAA 261448  
 900 sSerGlnGlnGluLeu...ValLeuLeuAlaGlnGluLysGlnGlnLeu 916  
 261449 ACGGTGCAAGATTTCAAAACCTTGAGGCTAACCAAAAACAATTCAT 261498  
 916 LeuValGluLysLysLeuGlnGlnSerLeuAspProLeuSerLysGlu 932  
 261499 ATCAAAATCAAGCTTACTTGAAGTGAACCTGATAACTTAACAGAGA 261548  
 933 LysGlnSerLeuLeu ..... GlnGluTyrAsnH 942  
 261549 AACCAAGCTTTTGAACCTTACGTAAAGAAACAACATGAGAGGTTGATGC 261598  
 942 AleuLysGlnLysLeuAspGlnGluTyrHisGlnLeuAlaGluArgLys 959  
 261599 TATTAACAAAGGTTGAGTATTAACATCAAGCTTTAAATATGCAACAG 261648  
 959 rGluPheGlnGlnGlnLeuAspAlaLeuGlyArg ..... LeuAsnMet 973  
 261649 CAGAGCTAGATAGAAAAACCCATGAGTTAAATATGCTTTTAAACCAT 261698  
 974 LysIle ..... LysGlyTyrLeuAspSer ..... LysLysAs 984  
 261699 GATCGGATCAAAAAGATCTTACAGCAACCACTGACAACTGTTAAAGAGC 261748  
 984 nGluLysLeuLysGlnLeuGlnGlnLysArgHisValLeuCysHisSerGln 1001  
 261749 CCAAAAATTAATGATTTAGAA...CGTACGCACTGCTTGAAGAACAC 261795  
 1001 euGln ..... SerCysMet 1005  
 261796 GTGAGTTGCTGAAGATGTTGCTGTTTAAAGCCCATGCTGTAATAA 261845  
 1006 AlaLysGlnLysArgIleSerAlaGlnLeuAsnLysSerLysGlnLeu 1022



alignment\_scores:  
 Quality: 436.50 Length: 1111  
 Ratio: 0.751 Gaps: 48  
 Percent Similarity: 52.295 Percent Identity: 20.612

alignment\_block:  
 US-09-538-396-2 x AAQ10378 ..

Align seg 1/1 to: AAQ10378 from: 1 to: 5858

```

215 GlnThrValLysAspGlnAlaHisLysLeuArgGluAsnIleAlaGln.. 230
      |||.....|.....|.....|.....|.....|.....|.....|
1196 GAGGCCCTGAAAGAGAGACAGACACATTTGACAGCTGCTGCGGAAAG 1245
      |||.....|.....|.....|.....|.....|.....|.....|
231 AspGlnLysSerAspAlaSerLys.....SerGlnMetGluGlnL 245
      |||.....|.....|.....|.....|.....|.....|.....|
1246 GGATCTGGAGAGGGCGGAGGTGCCAAGGCCACGACGCGGGGAGAA 1295
      |||.....|.....|.....|.....|.....|.....|.....|
245 euLysGluLysIleCys.....GlyThrGluArgGluIleLeu 257
      |||.....|.....|.....|.....|.....|.....|.....|
1296 TAGACAGAGAGCTAGCTGCGCCGCGGAGCGACATGACCGACATGCTCTG 1345
      |||.....|.....|.....|.....|.....|.....|.....|
258 GlnMetGluThrSerLeuAspGlnLeuArgArgLeuGlnGlyGlnIleAs 274
      |||.....|.....|.....|.....|.....|.....|.....|
1346 GAATTGGAAAGCCAAATGGACAGCTGCGACACATGCTGGAAGCTGCTGA 1395
      |||.....|.....|.....|.....|.....|.....|.....|
274 pIleLysAlaThrGluArgSerThrLeuLeuThrGlnGlnHisGluLysL 291
      |||.....|.....|.....|.....|.....|.....|.....|
1396 CAGGGAGAAAGGTGGACCTTCTCAACACAGCTTGAAAGAGAAAGAAAG 1445
      |||.....|.....|.....|.....|.....|.....|.....|
291 euAlaAlaLeuSer.....GluGluAsnGluAspThrAspGluGlu 304
      |||.....|.....|.....|.....|.....|.....|.....|
1446 TTGAGAGACCTTACGTTCCGCGTTGAGAAAGAAATCAATTACCAAGGTGAT 1495
      |||.....|.....|.....|.....|.....|.....|.....|
305 LeuMetGluThrGlnThrLysPheGluGlu..ArgIleAlaLeuGluG 320
      |||.....|.....|.....|.....|.....|.....|.....|
1496 CTT..GAGACGCGAGCCAAACTGAGAGCATGCCGCAATTAGAGAGCTTGA 1542
      |||.....|.....|.....|.....|.....|.....|.....|
320 u.....ThrLysIleSerLysLeuValArgAsp 330
      |||.....|.....|.....|.....|.....|.....|.....|
1543 ACAGAGCCCTGCTTGTGAAAGACCAAGCTGCAAACTCCAGAGGAGT 1592
      |||.....|.....|.....|.....|.....|.....|.....|
330 etaAspAsp.....GluAlaSerTyrSerSerVal 339
      |||.....|.....|.....|.....|.....|.....|.....|
1593 TAGAAGACACTAGGTGCTGCTACAGTTTCAGAAAGATCACCTATATAGAA 1642
      |||.....|.....|.....|.....|.....|.....|.....|
340 LeuSerLysGlnAsnSerGluLeuThrHisGluIleGlyLysLeuGln 355
      |||.....|.....|.....|.....|.....|.....|.....|
1643 CTGGAGAAAGACCTAGCATTTGAGATGACAGAGTAGCTAGCTCCGAG 1692
      |||.....|.....|.....|.....|.....|.....|.....|
356 .....AlaGluAlaAspAlaHisLeuThrM 364
      |||.....|.....|.....|.....|.....|.....|.....|
1693 AAGGCTAGAGCTCCATTAAGCTCTGTGGGATGTGACATCTCACTTCCC 1742
      |||.....|.....|.....|.....|.....|.....|.....|
364 euLysHisGluArgAspSer..... 370
      |||.....|.....|.....|.....|.....|.....|.....|
1743 TTTTGGACAGATAGCTTTTGCAGAAAGAAAGTTAGAACTCACCCCTACT 1792
      |||.....|.....|.....|.....|.....|.....|.....|
371 .....AspIleLysAsnIleCysThrLysHisAsnLeuGlyP 383
      |||.....|.....|.....|.....|.....|.....|.....|
1793 GACCACCCAGAGAGAAATTAATCTCTCTG.....AAGAGCATTTTGGAGC 1836
      |||.....|.....|.....|.....|.....|.....|.....|
383 ovalProGlu...HisProPheThrAsnAspValAlaMetAsnLeuThr 399
      |||.....|.....|.....|.....|.....|.....|.....|
1837 CCGGAGAAACTCATCAGAAAGAGATTAAGGCTCTATATACGCCACGG 1886
      |||.....|.....|.....|.....|.....|.....|.....|
399 snArgIleLysAlaArgLeuSerSerLeuGluAsnAspLeuLeuAspLys 415
      |||.....|.....|.....|.....|.....|.....|.....|
1887 AAAAGCTTCCAAAGACAGACAGTCATTGAAACCAAGCTGGACCATGCC 1936
      |||.....|.....|.....|.....|.....|.....|.....|
416 LysLysSerAsnGluAspGlnLeuAspValLeuThrPheLysHisTyrLeu 432
      |||.....|.....|.....|.....|.....|.....|.....|

```

```

1937 AACAAAGAACTCAGATGTGATA...GCTCTATGGAAGTCCAACTGGA 1983
432 sIle..... 433
1984 GACTGCATCGCATCCACACAGCCAGCGATGGAAGAACTGAAGATCTT 2033
434 .....AsnAlaArgTyrSerGluValAspGly 442
2034 TCAGCAAGGCGCTTGAAACAGACGCGCAGATTTGCTGAACCTAATAACA 2083
443 GlnIle.....GlnSerLysIleGlnSerMetSer 452
2084 CAATATAGAGAAATGAGACTAGATTACCAACAGAAATGAAAT..... 2128
452 rGlyIleLeuArgArgArgLysAspLysGluLysGluArgAspAlaAlaG 469
2129 .....TTGCGAAATCAACAAAGACTCTGACGCGCTCCCTGCTTAAG 2171
469 IuValGluLeuSerLysPheAsnLeuSerArgIleAspGluArgGluArg 485
2172 AGATGCAACCTTGAGGCGCTAACTGATGAAGTT...ATTAAAGAAAG 2218
486 HisMetGlnIleGluValGluArgLysThrLeuAlaLeuGlyGluArgAs 502
2219 GAAACAGCTGGAAGCCATCAGTGCAGAACTGGAACAACTGCAAGACCA 2268
502 pTyrAspSerIleIleSerGlnLysArgThrGluValTyrSerLeuGluG 519
2269 GCAT.....CTCGTAGAATGAGAACACGTTAAACAAATTCACAGAAAG 2312
519 IuLysIleLysValLeuLeuArgGluLysAspIleLeuAsnArgAsnAla 535
2313 CTGAATTAAGTA.....AAGGACCTAGAGTACTGCAAGCCAAATGC 2356
536 AspGluArgValLysLeuGlyLeuLys...LysAspAlaLeuGlnSerSe 551
2357 AATGAAACAAACCAAGGTATATGATTAATTTACATCAGACTCAAGGCTAC 2406
551 rLysAspLysLeuAsnGluLe.....ValAsnGluH 562
2407 TGAAGAAAGCTCTTGATGATCTTGTATGCTACCTCGAAGAACCAAGTCCGAAG 2456
562 IeLysAspLysIleLysLysValLeuArgGlyLysArgProPheGluLys 578
2457 GTAATTCGGAATGAAGAACTTAGACAGAGCTTGAGGAGCTGAGAA 2506
579 AspMetLysLysGluIleAsnGlnAlaPheTyrProValAspLysGluLys 595
2507 CAGATTAAACAT.....TTAGAGATTGAAAA 2532
595 rAsnGluLeuArgSerLysSerGlnGluAlaGlnGlnIleLysPheT 612
2533 GAATGCTGAAAGTAGCAAGGCTAGTAGCAATACCAAGAGAGCTCCAGGGGA 2582
612 hrgInserLysValThrAspAlaArgGluGlnLeuThrLysLeuArgArg 628
2583 GAGAGCTAAAGCTTACTAACCTTCAGAAATTTGATGAGAGCAGCA 2632
629 AspMetAspAlaLysArgArgPheLeuAspSerLysLeuGlnSerIleLe 645
2633 GTGAAGAGAGCTTGAAAGAAAGAACTTCAGATTTTGAAGAAAGATTGTC 2682
645 uGlnIleSerAlaAsnValAspMetPheProLysValLeuGlnAspAlaM 662
2683 TGAAGCTTCAGAGAGGAGCTCTCTGTTCAGAGAGAGTACAGAAACTG 2732
662 etaAsnLys.....ArgAspGluGln...LysArgLeuGluAsnPhe 674
2733 TAAATTAAGTACACCAAGAGAGAGACAGTTTAACATGCGTCTGTGAC 2782
675 AlaAsnGlyMetArgGluMetLeuAlaProPheGlnHisLeuAlaArg 691
2783 TTGAGAAAGCTGAGAGAAACTTAGCAGATATGAGAGCAAAATTTAGAGA 2832

```







[illegible]

332 spglnlaserlyserlyserleuauuileuulerlysglnasnsrglnleuthr 348  
1380 AACTAATATTAACAACTGAGCTTCATCAATCGAAGAAATTCAGAACTGCT 1429  
349 Hlsglnllegllyleuenglalnlaaglnlaaspaahlsleuthrmely 365  
1430 GTTGAACAGAGGCTGCTGTACAGCTGCAGACAGATGCCATTCAGAAACATAT 1479  
365 shsglnuargaspseraspilleylsasnilecysThrlyshlsasnleug 382  
1480 CCGAGCTAGAGATTCATTATATCACTCTTTGGCAACACAGCTAGAAATTTGG 1529  
382 lYProvalProclunhspropheThrasnaspyalaAlamelaslnleuthr 398  
1530 ATGGCTTTGAGGCTGGACCTTCACCTCACTGGAAGAAGCATTTAAATTTTTCAC 1579  
399 AsnArglilelyAlaArg.....LeuSerSertle 408  
1580 AAACCTTGTAGAGAGACAGACAAAGGCGAAGCAAAACTGCCCAACCACT 1629  
408 uglunasAspleuauAspylylslysSerasnglunspgluleuhsrv 425  
1630 GATGATGACTCTTGCGAAGAAAGAGACTCTGAAACAAAGACAGATAGAT. 1678  
425 alLeuthrplyshslytleuylsIleasnAlaArgTysSerGlunAlasp 441  
1679 .....GAGATTAAGATTAACAAAGAACTGGACTGGGA 1708  
442 glylnlleglnserlyslleglnsermetserylileuArgarg 458  
1709 AGAATTAATGAGTTAAATCAACAAATCCTAAT.....AA 1743  
458 glysAspylysglnuargAspAlaAlaaglunaleuLeuSerlysp 475  
1744 GAAGCAGAAATGAGCTCAAAATGTGAAGTATGATTAACAGCAGTTGGAG 1793  
475 heasnleuSeratgylleAspglunargglunarghismecgluleuAl 491  
1794 GATCTTCAGACAGATCTTGTGAACGTGGACAGAGACTC...ATTAAGACT 1840  
492 glunArglysrThrleuAlaleuenglunargpyrAspserIleIleSe 508  
1841 GAAACGTGAG.....TTAAGCAAGCTGTGAAAGAAACAGCAATGTGCA 1881  
508 rglnlysrArgThrGlunAlaTysSerleuenglunlys..... 520  
1882 AACCTTAATAAATGAGATTAAGTCTCCAAATGAAGAAAGACAGACTTGA 1931  
521 .....IlelysvAlleuLeuargglunlyAspIleIleasnArgasn 534  
1932 ACAGAGACCCGTGCTAAATCTTGACAGAGAGATGGAGAGTTTAACCATCAT 1981  
535 AlAspoglunArgAllylsleugly..LeuylslyAspAlaLeuGlSe 550  
1982 ACAACAACACAGTACCCAAATGAGATGCTGCACCAAGACAAAGCTGACAA 2031  
550 rtserylAspLyLeuasnglulleyalasnglunhlslyAspIlyIlel 567  
2032 AGAT..GAACAAATCAGAAAAATATAAATCTAGGCACAGTGTATGAATTA 2078  
567 yslsylvAlleuarglyArgasnProphcglunlyAspMetlyslsclu 583  
2079 CCTCACAGCTTG...GGATATTTTCCCAACAAAAACAGCTTGACAC... 2122  
584 IleasnclnAlaheThrProvalAspIlysglutyArgasnGluleuArgSe 600  
2123 .....TGG.....CTACATAG 2133  
600 rlysserGlnGlualaglunleuylspherThrglnserlyValT 617  
2134 TAAATCAAAAGAAATTAATACAG..... 2155  
617 hrAspAlaArgglunleuthrlylsleuArgArgAspMet..... 630

```
2156 .....ACGAGGACAGACTTGCCTTGAATTCACAGAGACTTACATCATCT 2200
631 AspaLalysArgArgPheLeuAspSerLeu...GlnSerIleLeuG1 646
2201 GAGCAGAAATAAATCATATAATATAATGAAATGAAAGAAAGAGAGCA 2250
646 nIleSerIlaAsnValAspMetPheProLysValLeu.....GlnAspA 661
2251 GTTGTCCAGTTACGAAAGCAAGCTGTGTGATGTGTGTGTGTGTGTGTGT 2300
661 lAmEaSnLysArgAspGluGlnLysArgLeuGlnAsnPhaLlaAsnGly 677
2301 TTGAAAGTGAATTTAGACAGCTTAAAGAGAAATGAAATGAAATCATCAAA 2350
678 MetArgGluMetLeuAlaPro.....PheG1 686
2351 CAGCAGAGCATGCTGCTGAGCAGCAAGCATTTACTCCAGTTCATTCAC 2400
686 uHlSerLeuAlaArgLysAsnHis...ValCysProCysGluArgAlaP 702
2401 TCAGCTAACAGAGCAAAACAGCTCATGTGTGTGTGTGTGTGTGTGTGTGT 2450
702 he...ThrProAspGluGlnAspGluPheValLysGlnArgMetGln 717
2451 TTCACACAGAGGCTGAGTTACAGAAAGTCATCAGTGATTTGCGCTTAA 2500
718 AsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSerAs 734
2501 CTGCGACATTGCTCCAGATTAATCAAGTCAACAGATTCAGAGCTAAATA 2550
734 nAlaGluAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyAspA 751
2551 AAAGCAAAAGCGCGTGATGAATGCTGTGTGTGTGTGTGTGTGTGTGTGT 2600
751 lArgValLysLeuValGluGlnThrIlePro...LeuAlaGluLys... 765
2601 GCATTAATGATTGTAAGAGAGAAATACCAAGATTAAGAAACAACGTCG 2650
766 ...AspLeuAsnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaP 781
2651 CAGATTCGATATAGACACATACAGCCCTAAAGAACAGACATAGAAACA 2700
781 eaSpAspLeuLeuGly..... 786
2701 AGAAACACTCTGGTACATAATAGCCTGAGAGAAAGTGCCTAAAGTAT 2750
787 .....ValLeuAlaHisValGlnMetAspArgAla 797
2751 GCCTGACAGATGTTACAATATATGAGAGGTTCCAGATGGAATTAAGAT 2800
798 ValGlu.....AlaLeuLeuGlnProThrAspTh 807
2801 GTTGAAGAAAATTCACACAGACAGCTAACTCAAGAGATATAC... 2848
807 rIleAspArgHisValHisGluIleGlnGlnLeuValLysGluValGluA 824
2849 ..TTAGATCGAAGCTGCCAACAGTCAACAG.....GAGAAACAAG 2888
824 sPLeuGlnUryAlaLeuAspSerSerGlyArgGlyValLysSerLeuGln 840
2889 AGAAACAGCAAGTTAGACACA.....GTTCTCTAGTAAGATT 2926
841 GlnIleGlnLeuGlnUryLeuAsnPhaLeuGlnArgThrArgAspThrLeuI 857
2927 GAATTCATCGTAAGCTT..... 2944
857 eValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGluAspM 874
2945 .....ATACAGACACAGCAAGCAAGATTCACACATTA 2978
874 eCSeSerAlaGlnValArgThrHisAsnAlaArgGluGlnUryLysValLys 890
|||||.....
```

```
2979 AAAGTACACA.....AATGACATAAAATCGAAGACTTCAG 3016
891 AlaseSerIleLeuGlnUryPheGlnLysSerGlnGlnGlnValLe 907
3017 AATACCACTAATTTGCAAGCTGTCAGCAACTGAGGAGACACAGCTGGA 3066
907 uLeuAlaGlnGlnUryGlnGlnLeuIleValGlnUryLysLeuGlnG 924
3067 ATTATCCAGTGAAGTTCAGTCTTTGTACAGAGATTAAGATGCTTAAG 3116
924 lUSeLeuAspProLeu.....SerLysGluLys 933
3117 ACAGGTAAAGCCCTTTGAAACAACATGGAAGAGTTCACAGCAAGAAAA 3166
934 GlUSeLeuLeuGlnGlnUryHisAlaLeuLysGlnLysLeuAspGlnG 950
3167 GAAGATTAATCAACAAAAAATGACACCAAAATAGCAGCAAGTGA 3216
950 uUryHisGlnLeuAlaGlnUryArgLysArgLysPheGlnGlnUryLeuAspA 967
3217 ACTGAATGATTAATAAGAGAGGTTAAATATCTAGGCTATAG... 3262
967 lAlaLeuGlnArgLeuAsnMetLysIleLysGlyTyLeuAspSerLys 983
3263 .....AAAGCATTCAGAAATTAATTCAGATGGGAA 3295
984 AsnGln.....LysLeuLysGlnLeuGlnGlnUryArgHisValLeuCyH1 998
3296 GACAGCTATAGACCAAAAGAACTGAATTAATTAATTA... 3340
998 sSerGlnLeuGlnSerCyMetAlaLysGlnGlnArgIleSerAlaGlnL 1015
3341 .GCTCACTAAGTAATGCGAGAAACACAAAGAAAGATTAATGAAGATA 3389
1015 euAsnLysSerLysGlnLeuLeuGlnGlnGlnGlnUryLysArgAsn 1031
3390 TCGAGACTCATGACACAAGATTTGATACACAGAAAGTACAGAAAGCTGG 3439
1032 lLeAspAspAsnLeuLysTyArgLysThrLysAlaAspValGlnGlnLe 1048
3440 CTCAAGATTAACCTTACTTAAGAAAGAAATGAGAACTTAAGAAAGT 3489
1048 uThrArgAspIleGlnSerLeuGlnUryArgLeuLeuSerIleGlySerI 1065
3490 TGA.....GAAGAAAGAAACACATTTGAAGAAATGGCGTAA 3530
1065 euSerAlaIleGlnAlaAspLeuLysArgHisSerGlnLys..... 1079
3531 TCGAGGTTTGCATGAAGAAAGTGAACATCAGAACTGGAGAGACATA 3580
1080 GlUArgLeuAsnSerGluPheAsnArgTyArgGlnGlnUryLeuSerValTy 1096
3581 GACAAATATAAAGAAATCATATTTGGCATTTGGCGCAGACAGAAAGTTA 3630
1096 rGlnSerAsnIleSerLysHisLysGlnGlnUryLeuLysLeuSerGlnTyL 1113
3631 TGAAGAGAAATTAATTCATTTAAGAAAGAACTCGAAGAACCAATTTTC 3680
1113 yAspIleGlnUryArgTyThrAsnGlnPheLeuGlnLeuLysTyThrH 1129
3681 GCGATGCTGAGAAAGATATAGAAATATGATGATTTTATGAGAGACACA 3730
1130 GlUmetAlaAsnLysAspLeuAspArgTyTyThrAlaLeuAspLysAl 1146
3731 GAAGCTGTGAACAAGATCTGATATTTAATAAGACCTTGACCAAC 3780
1146 AluMetArgPheHisSerMetLysMetGlnUryLysLeuLysIleLeI 1163
3781 AATTAATGAATTTCAACAGATGAATGAAGAAATCAATTAATTAATAC 3830
1163 yGlnUryLeuTyArgGlnUryArgGlnGlnAspIleAspTyIleSer 1179
3831 GTGACCTGTGGCAGAGTACCTATCTGACACAAATATGATATCAATAGAA 3880
```



111 |||..... |||..... |||.....  
980 ACACAGGCTAGAACTAAAGATATCAATGGAAGTAAATATCTGAA 1029  
215 nThrValLysArgGlnAlaHisLysLeuArgGluAsnIleAlaGlnAsp 232  
1030 GCATATGAAAGAAAAGCTTGATGATCTGATCAGATTACAACTAAG 1079  
232 InGluLysSerAspAlaSerLysSerGlnMetGlnInLeuLysGlu 248  
1080 AAGCCCAATTAACATCTTCAAGGAATTCCTAATCTATGAGATGAA 1129  
249 IleCysGluThrGluArgGluIleLeuGlnMetGluThrSerLeuAsp 265  
1130 CTGTGATCCATTGAAAGATCTGTAAGAAATGAAATTAATCTCTTA 1179  
265 uLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSer 282  
1180 AATATATGAACCTTGCAATGAATTAAGCTTGATAGCCGGAAGAGC 1229  
282 hLeuLeuThrGlnGlnHisGluLysLeuAlaAlaLeuSerGluAsn 298  
1230 AATGAGAGAAATATATAGTGAAGTGAAGAGAAATGGAAGCTTTT 1279  
299 GluAspThrAspGluGluLeuMetGluTrpGlnThrLysPheGluArg 315  
1280 CAGGAGCTGATGACCACTAAATGACTTATATCACAATCACCAGAAC 1329  
315 GluAlaLeuLeuGluThrLysIleSerLysLeuValArgAspMetAsp 332  
1330 AGTAAGGAGAAAGAAAGAAATGTGATGACTCTCATCTGAACTGAAA 1379  
332 spGluLysSerLysSerValLeuSerLysGlnAsnSerGluLeuThr 348  
1380 AACTAATATGAATCTAGCTCTCAATCAGGAAATCGAATCTGCTT 1429  
349 HisGluIleGluLysLeuGlnAlaGluAlaAspAlaHisLeuThrMet 365  
1430 GTTGAAACAGGCTGCTACAGCTGCAAGCAGATCGCCATCAGAACAT 1479  
365 sHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeu 382  
1480 CCGAGCTAGAGATTCATTAATTCAGCTTGCGCACACAGCTGAATGG 1529  
382 LysProValProGluHisProPheThrAsnAspValAlaMetAsnLeuThr 398  
1530 ATGGCTTGAGCGTGGCCATTCAGTGAAGACAGATTAATAATTTCAC 1579  
399 AsnArgIleLysAlaArg.....LeuSerSerLe 408  
1580 AAACCTGTAGAGAGAGACAAAGGGAAGCAAAAACCTGCCAACCACT 1629  
408 uGluAsnAspLeuLeuAspLysLysSerAsnGlnLysGlnLeuAsp 425  
1630 GATGAATGACTTTCAGAAAAGAGACTGCAACAAACAAACACATGAT 1678  
425 aLeuThrLysHisLysLysIleAsnAlaArgTyrSerGluValAsp 441  
1679 .....GAGATAAGAGATAGAAAACCTGAGCTGGA 1708  
442 GlyGlnIleGlnSerLysIleGluSerMetSerGlyIleLeuArgArg 458  
1709 AGAATATGACTTAAATCAGAAATCCTAAGT.....AA 1743  
458 GlyAspLysGluLysGluArgAspAlaAlaGluValGluLeuSerLys 475  
1744 GAAAGCAAGATGAGCTGAAAAATGTGAAGTATTAATACACAGTGGAG 1793  
475 hAsnLeuSerArgIleAspGluArgGluArgHisMetGlnIleGluVal 491  
1794 GATCTTCAGACAGATCTTGAAGCTGACAGAGAGCTC...ATGAAGCT 1840  
492 GluArgLysThrLeuAlaLeuGluArgAspTyrAspSerIleIleSe 508  
|||||..... |||..... |||.....

1841 GAAGCTGAG.....TTAAGCAAGCTGAGAAAAACGCAATGTAGA 1881  
508 rGlnLysArgThrGluValLysSerLeuGluGlnLys..... 520  
1882 AACCTTAATAATGAAATGAATTAAGTCTCAAAATGAAAAGCAGATTAG 1931  
521 .....IleLysValLeuLeuArgGluLysAspIleIleAsnArgAsn 534  
1932 ACAGAGACCTGCTTAACCTGACACAGAGATGAGAGAGTAAACATCAT 1981  
535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGlnSe 550  
1982 ACAACAACAGCTACCAAAATGAGATGCTGACCAAGACAAACCTACAA 2031  
550 rSerLysAspLysLeuAsnGluIleValAsnGluHisLysAspLysIle 567  
2032 AGAT...GAACAATCAAGAAAAATTAATTAATGACACAGATGAATTA 2078  
567 yLysValLeuArgGluArgAsnProPheGluLysAspMetLysGlu 583  
2079 CCTCAGCTGTG...GGATATTTCCACAAACAAAAGAGCTGAAGAC... 2122  
584 IleAsnGlnAlaPheTrpProValAspLysGluTyrAsnGluLeuArgSe 600  
2123 .....TGG.....CTACATAG 2133  
600 rLysSerGlnGluAlaGluGlnLeuLysPheThrGlnSerLysVal 617  
2134 TAAATCAAAAAGAAATTAATCAG..... 2155  
617 hAspAlaArgGluGlnLeuThrLysLeuArgArgAspMet..... 630  
2156 .....KCCAGGAGACAGCTTGCCAAATTGAACAAGAACTAGCTTATCT 2200  
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuGln 646  
2201 GAGAGAAATTAATAATCATTAATTAATGAATCAAAAAGAGGAAGAGCA 2250  
646 nIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAsp 661  
2251 GTTGCTCAGGTAGCAGACAGCAAGCTGTGATGTGTTGTGAGCAGAGATT 2300  
661 lMetAsnLysArgAspGluGlnLysArgLeuGlnLysAsnPheAlaAsnGly 677  
2301 TTGAAGTGAATTAAGACAGCTTAAGAGAAATGGAATAATCATCAAAA 2350  
678 MetArgLysMetLeuAlaPro.....PheGln 686  
2351 CAGGAGCCATGCTGCTGAGAGCCACAGCAAGTTTACTCCAGTTCAATTAC 2400  
686 uHisLeuAlaArgLysAsnHis...ValCysProCysGlyGluArgAla 702  
2401 TCAGCTTAACAGACGAAACAGATCATGTGCTGCCGTGTCAGAGAGATT 2450  
702 hE...ThrProAspGluGluAspGluPheValLysLysGlnArgMetGln 717  
2451 TTCAGACAGAGGCTGAGTTACAAAGAGCTCATGATGATTGCAAGCTAAA 2500  
718 AsnSerSerThrAlaArgSerLysAlaLeuAlaMetGluSerSerAs 734  
2501 CTGCGACTTGCTCCAGATTAACCTCAACTCAACAGAAATCAAGAGCTAAAA 2550  
734 nAlaGluAlaLeuPheGlnGlnLeuAspLysLysLeuArgThrIleTyrAsp 751  
2551 AAAGAAAGAGCGGCTGATGAATGCTGGGCTGTGCCATGAGGCAAA 2600  
751 lArgValLysLeuValGluGluThrIlePro...LeuAlaGluLys... 765  
2601 GCATTAATGATTGAAGAGAGAAATACCAATTAAGAAACAAACCTG 2650  
766 ...AsnLeuAsnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaPhe 781  
2651 CAGAAATGCTCAATTAAGACATACAGCGCTAAAGAAAGACATAGAAAGACA 2700  
|||||..... |||..... |||.....

781 eaaspbleuengly..... 786  
7701 AGAAGACCTCTGGTACAAATATGCTGAGAGAAAGTCCCAAGTAT 2750  
787 ..... ValLeuAlaHisValGlnMetAspArgAla 797  
2751 GCCTGACAGATGTTACATATATGAGAGGTTCCAGATGGAATTAAGAT 2800  
798 ValGlu..... AlaLeuGlnProThrAspThr 807  
2801 GTTGAAGAAAAATGCACAAACAGCAGCTAGCTACAGGAATGAC.. 2848  
807 rilaasparghisvalhisglnileglnleuvallysgluvalglua 824  
2849 ..TTAGATGGAAGCTCCCAACAGTCAACAG..... GAGAAACAG 2888  
824 srpleuglutyAlaLeuAspSerSerGlyArgGlyVallySerLeuGlu 840  
2889 AGAAACAGCACAGTACAGTACACA..... GTTCTAGTAGATT 2926  
841 GlnileglnleuglnleuAsnPhelLeuGlnArgThrArgAspThrLeuI 857  
2927 GAATTGAATCTTAAGCTT..... 2944  
857 evalgluvalAspAspLeuArgAspGlnHisArgMetLeuAsnGluAsp 874  
2945 .....ATACAGACAGCAGCAGACAGATCAACATCA 2978  
874 etserSerAlaGlnValArgThrHisAsnAlaArgGlnLysVallyS 890  
2979 AAGATACACA.....AATGAGCTAAATCTGAGAACTTCAG 3016  
891 AlaSerSerIleLeuGlnArgPheGlnLysSerGlnGlnGluLeuVal 907  
3017 ATATCCACTAATTTCACACGTCGTCAGCACTGAGAGCAGACGCTGA 3066  
907 uLeuAlaGlnLysGlnGlnLeuIleValGlnLysLysLeuGlnG 924  
3067 ATTATCCACTGAGTTCAGCTCTTGTACAGAGATAAAGATGCTTAAG 3116  
924 luserLeuAspProLeu.....SerLysGlnLys 933  
3117 AGCAGGTAGCCCTTGGAAACACATTCGAAAGTCCACGCAAAAAA 3166  
934 GlnuserLeuGlnGlnLysAsnAlaLeuLysGlnLysLeuAspGln 950  
3167 GAAGATTAATCAACAAAAAATACAGCAACAAATAGCAGAGATTA 3216  
950 utyrHisGlnLeuAlaGlnArgLysArgLysPheGlnGlnLeuAsp 967  
3217 ACTGATATGATTAAGAGAGTAAATATTCATGCTATAG... 3262  
967 lAleuGlnArgLeuAsnMetLysIleLysGlyTyrLeuAspSerLys 983  
3263 .....AAGCATTGAGAAATTAATTCAGATGCGAA 3295  
984 AsnGlu.....LysLeuLysGlnLeuGlnGlnLysValLeuCyH 998  
3296 GACGACTATAGACCAAAAGAACTGAATTAATAAGTATA... 3340  
998 sSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgIleSerAlaGln 1015  
3341 ..GCTCACTAGTGAATGAGAGAACACAAAGAAATTAATGAAGATA 3389  
1015 euAsnLysSerLysGlnLeuGlnGlnGlnGlnLysArgAsn 1031  
3390 TGAGACTCATGACAGAGATATTGATACACAGATACCAAGAAAGGTGG 3439  
1032 lIleAspAspAsnLeuLysTyrArgLysThrLysAlaAspValGlnLe 1048  
3440 CTACAGATTAACCTTACTTAAGAAAAAGAAATGAGAACTTAAGAACT 3489

1048 uThrArgAspIleGlnuserLeuGlnArgLeuLeuSerIleGlySerL 1065  
3490 TGA.....GACAAAGAAACACACATTTCAGAGAAATGGCTCAA 3530  
1065 euSerAlaIleGlnAlaAspLeuLysArgHisSerGlnLys..... 1079  
3531 TGCAGGTTTGCAAAATGAAAGTAGACATCAGAGTGTGAAGAGACATA 3580  
1080 GlnArgLeuAsnSerGlnPheAsnArgTyrPheGlnGlnLysSerVal 1096  
3581 GACATATTAAGAAAGAAATTCATATTTGCCATTAGCGACAGAAAGCTTA 3630  
1096 rGlnSerAsnIleSerLysHisLysGlnGlnLeuLysSerGlnTyrL 1113  
1113 ysaspileglnLysArgTyrThrAsnGlnPheLeuGlnLysThrThr 1129  
3681 GCGATGCTGAGAGAAAGATAGAGAAATGATGATTGTATGAGAGACAA 3730  
1130 GlnMetAlaAsnLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAl 1146  
3731 GAACCTTGAAACAAGATCTGATATTATTAAGACTCTTGACCAAGC 3780  
1146 AleuMetArgPheHisSerMetLysMetGlnGlnLysLysIleIleL 1163  
3781 AATATGAAATTTCCACCTATGAAATGGAAGAAATCAATAAATTAAC 3830  
1163 ysGlnLeuTyrPheGlnThrTyrArgGlnAspIleAspTyrIleSer 1179  
3831 GTGACCTGTGCGAAGTACCTATGTCGACAAAGATATGATACATAGA 3880  
1180 lIleAsnSerAsp.....SerGlnGlnAlaGlnTyrArgSe 1191  
3881 ATACGCTGTGATGCCGAGAAATGATATACCTCTGATAAGCGGAA 3930  
1191 rTyrSerTyrArgValAlaMetGlnThrGlyAspAlaGlnLeuGlnMetAl 1208  
3931 TTATTAATCAACGAGGTGATGCTGAAGGAGACACACCTTGGATATGC 3980  
1208 rGlnArgCysSerAlaGlnLysValLeuAlaSerLeuIleLeuArg 1224  
3981 GAGAGCATGACAGTCTGACAAAGATATAGCTCATCATCTGC 4030  
1225 LeuAlaLeuAlaGlnThrPheCysLeuAsnCySGlyIleLeuAlaLeuAs 1241  
4031 CTGGCCTGCGTGAAGCGTCTGCTCAACTGTGACATCATTCCTTGA 4080  
1241 pGlnProThrThrAsnLeuAspGlyProAsnAlaGlnuserLeuAlaAla 1258  
4081 TGAGCCAACCAAAATCTTGACCGAGAAACATTAATCTCTTGACATG 4130  
1258 lAleuLeuArgIleMetGlnAlaArgLysGlnGlnLysAsnPheGlnLeu 1274  
4131 CTCTGCTGAGATATTAAGAGTCCCTCAGACAGCAGTAACCTCCAGCTT 4180  
1275 lIleValIleThrHisAspGlnArgPheAlaHisLeuIleGlnArgGln 1291  
4181 CTGATATCACTCATGATGAAGATTTGTGGAGCTTTTAGAGCGCTTGA 4230  
1291 GlnLeuAlaGlnLysTyrTyrArgValSerLysAspGlnGlnHisSerI 1308  
4231 ATATGTGAGAAATTTCTACAGGATTAAGAAAGAAATCGATAGTCTCAG 4280  
1308 lIleLeu 1309  
4281 AGATT 4285

seq\_name: /cgn2\_6/protdata/2/lna/5a\_COMB.seq:US-08-328-254-5

seq\_documentation\_block:

; Sequence 5, Application US/08328254

; Patent No. 5710022

GENERAL INFORMATION:  
APPLICANT: Zhu, Xueliang  
APPLICANT: Lee, Wen-Hwa  
TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/328,254  
FILING DATE: 24-OCT-1994  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/141,239  
FILING DATE: 22-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CU 1191  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8789 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 544..7990  
US-08-328-254-5

alignment\_scores:  
Quality: 439.50 Length: 1256  
Ratio: 0.668 Gaps: 49  
Percent Similarity: 52.389 Percent Identity: 19.904

alignment\_block:  
US-09-538-396-2 x US-08-328-254-5 ..

Align seg 1/1 to: US-08-328-254-5 from: 1 to: 8789

51 CysThrGluGluLeuProAsnSerArgLysIsthrPheValH1 67  
||||| |||:  
3808 TGCCTGGGAAACAGCTCCCGACATACCAAT.....TA 3839  
67 SASPrOLysValAlaGluGluThrGluThrLysGlyGlnIleLysLeuA 84  
||||| |||:  
3840 TGACCT.....CCAGGGGAAGATAAACCAGGGCTTCAGAA.... 3879  
84 rgrPheLysThrAlaAlaGlyLysAspValValCysIleArg.....Ser 98  
||||| |||:  
3880 .....TGCAATTCGAATGCA 3897  
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPhe.....Lys 111  
||||| |||:  
3898 TTTTCTGTCTCAATGCTTGTGATGATTCCTCGGGAATCAGCA 3947  
111 SAAlaIleGluSerValLeuGlnThrIleAsnProHisThrGlyGluLysV 128  
||||| |||:  
3948 AGATATCATTAATCTCAACTGCGGGTAAAGAGACATCAATGGAAT 3997

128 aLysLeuSerThrArgCysAlaAspMetAspArgGluIleProAlaLeu 144  
::: ||| ::| |||:  
3998 TGAGTTACTTCTCATGTGATAGAGACCGTGCAGCAAGAAAGTTGAAGTTG 4047  
145 MetGlyVal.....SerLysAlaValLeuGluAsnValI1 156  
||||| |||:  
4048 CTAAATGAATGAAGAATTAAGTACCTCAAACTCCATTACAGAGGTCA 4097  
156 ePheValHisGlnAspGluSerAsnTrpProLeuGlnAsp...ProSer 172  
||||| |||:  
4098 ACTATGACCAAAATTTGAAGCATGATGAAATTTGAAAAAATGTTGGG 4147  
172 hLeuLysLysPheAspAspIlePheSerAlaThrArgTyrThrLys 188  
||||| |||:  
4148 AACTTAAGAAAGAAACATCATTTA.....AGTGA 4179  
189 AlaLeuGluValIleLysLysLeuHisLysAspGlnMetGlnIleLys 205  
||||| |||:  
4180 AAATGGAATATTTTCTGTGATCCAGAGTACTCCAGAGATGA 4229  
205 sthrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspGlnAlaH 222  
||||| |||:  
4220 AACTTGTGAGGCCCTCATTTCTGATTGAAATGATCAGATTAATCAT 4279  
222 IsLysLeuArgGluAsnIleAlaGlnAspGlnLysSerAspAlaSer 238  
||||| |||:  
4280 CA.....CGTAAGATATGAGATTAATGCGCCAAAGTAATGACAGC 4323  
239 .....LysSerGlnMetGluGlnLeuLysG1 247  
4324 TGGAAAGAGATTTCTGTGATGCGAAATGAGCTGCAATGATGATGATC 4373  
247 uLysIleCysGlyThrGluArgLuiLeuGlnMetGluThrSerLeuA 264  
||||| |||:  
4374 GAGAAAGCTAGCATTTGAGCATGAAAGCCCTCAGCTGAGGCTGACTAG 4423  
264 sPgluLeuArgArg..... 268  
4424 AGGTATGTCAAACAGAGAACTATGTTAGAAAAAGACATGAATGAAG 4473  
269 .....LeuGlnGlyGlnIleAspIleLysAlaThrG1 279  
4474 CAGAAAGTATTTCTGCTGCTGAAAGAGACCTCAGTGCACAGAGCA 4523  
279 uArgSerThrLeu.....LeuThrGlnGlnHisGlu 290  
||||| |||:  
4524 GAGAAACAGCTCGTGGAAATTAAGATACATGTCATAAAAAACACAGG 4573  
290 yLeuAlaAlaLeuSerGluGluAsnGluAspThrAspGluLeuMet 306  
||||| |||:  
4574 CACTGATCAGTGTCTGAAATAATGAAGAGAAACAGAGAGCTTAG 4623  
307 GluTrpGlnThrLysPheGluGluArgIleAlaLeuLeuGluThrLysI1 323  
||||| |||:  
4624 TCTCATCAAGAGAGTGTCTCATTCATTCAGTGCAGAGGCGAGAGGT 4673  
323 eSerLysLeuValArgAspMetAspAspGluAlaSerTyrSerValI 340  
||||| |||:  
4674 GAAGGAAAGAGAGAGCTCTTCAAGACTTTGCTCTATGTGAGTGAGC 4723  
340 eSerLysGlnAsnSerGluLeuThrHisGluIleGlyLysLeuGlnAla 356  
||||| |||:  
4724 TGTAAAGACAAACATCATCTCCAGAAAGCTCAGAGTTGGAAG 4773  
357 GluAlaAspAlaHisLeuThrMetLysHisGluArgAspSerAspIleLys 373  
||||| |||:  
4774 GACTCAGAGGCTGCTTTGACAAATGATGAGTGAAGAAACCAATGTC 4823  
373 sAsnIleCysThrLysHisAsnLeuGlyProValProGluHisProPheT 390  
||||| |||:  
4824 ACACTGAATTAAGAG..... 4839



390 hransapvalalameasleuthrasnarlylleysalaargleuser 406  
4840 ..AAGGATTGCTTGTCAAGGAATCTGAAGCCTGCGAGCCAGAGCTAGT 4887  
407 Serleuglunansapleuleasplyslylseseransglunsglue 423  
4888 GAATCAGATATATGAAGGCTGAATGTCTCCAGGCCCTTGAGGCCGACT 4937  
423 uasapvalleutplyshistyleuyslleasnalargtyserglue 440  
4938 GGTGAGGAAGGATTCATTCATTCAGGCTGAGCTCAACAGAGAGGAG 4987  
440 alsapglunleuglinserylleuglusermetserglylleuarg 456  
4988 TGCAT...CAGCTGAGAGAGGCTGAGAACTGAGA.....GTTCGC 5028  
457 Argarglysaplysglulysgluargaspalaalaglulueuse 473  
5029 ATTGAGCCCATGAAGGAAGAGAGCTGCACATGCGAG... 5067  
473 rlyspheasleuserarglleaspluarargluarghismetglnleg 490  
5068 .....AACTGAAGAGACCGGAGCGG..... 5088  
490 luvalgluarlysthrleualaleucllygluaraspytyraspserile 506  
5089 .....GAGAAATGAT..... 5097  
507 lleserlunysargthrgluvaltyrserleuglunlunyslleysva 523  
5098 .....TCACTTAAGATTAAGTTGAGAA 5120  
523 lleuleargglulysaplleleasnarasnalaspgluargval 540  
5121 CCTGGAAGGGAATTCAGATGTCAGAAAGAAAC...CAGAGCTAAGTGA 5167  
540 yslleugllyleuyslyasapalaleucluserlyasplysleuasn 556  
5168 TTCTT.....GATGCC...GAGAAATTCAGAAAGCAAGAGTGAAG 5202  
557 glullevalansglunhlyasaplylleyleysvalleuarglylar 573  
5203 ACCTTA.....AAACACAAATAGAAAGAGTGCACAGAACCT 5240  
573 gasnprophleuglulysapmetlysglulleasnglnalaphetrrp 590  
5241 GAAAGTTTGAATTAAGCTTCGTC..... 5265  
590 rovalasplysglulysasngluleuargserlysserlunlualglu 606  
5366 .....ACGTTAAGGCTCTGMAAAAGAAATCTGACA 5295  
607 glunluleuyspnerhrinseryllevalthraspalarglunle 623  
5396 AAACAAATACAGAAACAAAGT.....CAGTT 5324  
623 uthrlyleuarghrasphetaspalalyasargphelueaspserl 640  
5325 GTCGAACTGACAAAGTACTCTTCATTTAAAGCTGTGAAGAA 5374  
640 yslleuglinserylleuglinserylleasnalaspmetpheprolys 656  
5375 AGGAGCAAGCAGAGATACAGATC..... 5397  
657 valleuglnaspalametasnlyasargaspglunlunysargleugluas 673  
5398 .....AAAGAAATCTTAAGCTGCAGTGCAGAT 5426  
673 npheliasnglymetarglunleuualaprophleuglnhileuualaa 690  
5427 GCTTCAGATATGATTAAGGCTAAATGAGGAGTAGAGGCTGTGTG 5476  
690 rglusasnhlivalcyaprocycygluaralaphetrrpaspglu 706

5477 GTGACCAAGAAATTAAGAGCCACAGACAGATGTAGACCCACCATATA 5526  
707 Gluasplunphevallylsyglunargmetglnasnserythrilaagl 723  
5527 GAGCAAGAG.....CATCAGCTGAGAAATGCTTGAAGAAAGCT 5564  
723 uargserlysalaleualametgluseraserasnalaaglualaleupheg 740  
5565 GAGAGCCCGCTTGAGAGCTGATGAAGAGAGAGCTC...TGCTGTAC 5611  
740 Inclunleuasplysleuarg.....Thrllyraspalatyryval 753  
5612 AACACTGAGAGAAAGTACAGCATCATCATTTACTTAAGGAGAGTGT 5661  
754 lysleuvalgluluthrlleproleualaglulysasleuasnlnh 770  
5662 GAGAACCTTGAAGAGAGCTGAGATGCGAGGCAAGC...CAGAGCA 5708  
770 sleualaspgluserglunlunysalaglnalaphasapleucllyv 787  
5709 TGCAGCTCTTGAGGCGAGAGATTCAGAA..... 5736  
787 alleualhnsivalglnmetasparaspalavalglualaleucln 803  
5737 .....GGAGAGTAGAGACCTTAAGCAAAATAGAGGAGTACCCAA 5781  
804 Prothraspthrilleasparghnsval..... 812  
5782 AGCTGAGAGCTCTGAAATTAAGATGTTGTACTATTAAGCTCAGAAAGAA 5831  
813 .....HISglulleuglunleuvallysgluvalgluaspleug 826  
5832 AAATTCGACAAATGAATTAACAAAGAGCAAGAGCGAATTCGAAATAG 5881  
826 lutyralaleuaspserseryllyarglyval.....lysleu 839  
5882 AA...ATAATTAATTCATTCATTTGAATAATTTGCAAGAAAGAACCA 5928  
840 glunlulleuglunleuasnphelueuglnargthraspthrle 856  
5929 GAGAAATTAAGATG.....AAACAAATTCAGAACAC 5960  
856 ullevalglulvaspasppluarasglnhlsargmetleuasnlnua 873  
5961 TGCATGAGATG.....CTTCAACACAAATTAAGAGCTCATAGAGA 6004  
873 spmetserasnalaglnalargtrphisasnalarggluglu...lys 888  
6005 GAGTGGCAGCCCTG.....CATATGACCAAGAAAGCTGTAG 6042  
889 vallylaserserylleugluargpheelunysserlunlunle 905  
6043 GCCAAGAGCAGAAAT..... 6057  
905 uvalleuleualaglululysglunlunleualglulysleul 922  
6058 .....CTTAGTAGTCAAGTAGAGTGTCTTAAGCTGAGAGGCTCACT 6100  
922 euglunluserleuasppro.....leuserlys 931  
6101 TGCACAAAGCCCTTGATGAGGCCCAAAATTAATTAATTTGTTTCAATCT 6150  
932 Glulysgluserleuenglunlutyrsanlaleuylsglunlunleuas 948  
6151 TCACTGATGAGCCCTCATTCAGAAAGTAGAGATGCAAGAGAGATGCA 6200  
948 pglunlutyrlhlsleuualaglualarglylsargglunpheelnglnlul 965  
6201 GAAGAGAGATGAAGAAATCACTAGACTGAAGAAATCAATTCAGAACCAAG 6250  
965 euaspalaleucllyargleuasnmetlyslleysglytyrleuaspsers 981

6251 AGCAGCTTGC..... 6261  
 982 LysLysAsnGluLysGluLeuGlnGlyArgHisValLeuCysHis 998  
 6262 .....TCTAACTGTCCTCCAGGTGGAAAGACACCACTTTGGAA 6302  
 998 sSerGlnLeuGlnSerCysMetAlaLysGlnArgIleSerAlaGlu 1015  
 6303 GGAGCA.....AAGTGAACAGTGAAGAAATCTGACACTGGAA 6340  
 1015 euAsnLysSerLysGluLeuLeuGlnGlyGlnGlnLeuLysArgAsn 1031  
 6341 TGAGCGAGAGATCCAGTGTCAATCAAA.....AATGCTCT 6381  
 1032 IleAspAspAsnLeuLys.....TyrArgLysThrLysAl 1043  
 6382 TTGACGACACATTAGAGTGTCCAGAGCTTCTTCAAGAAATCAAGAA 6431  
 1043 aaSPValGluGlnLeuThrArgAspIleGluSerLeuGluArgLeu 1060  
 6432 TGAGCTTGATTCACAAATGACAAATGCTCTTCTTGGAAAA... 6477  
 1060 euSerIleGlySerLeuSerAlaIleGluAlaAspLeuLysArgHisSer 1076  
 6478 .....GTAACAAATGACTGCAAGAACTGAGTGCAGAGGAATG 6522  
 1077 GlnGlu.....LysGluArgLeuAsnSerG 1085  
 6523 CATGATATGGCAGACAAACAGAGAGCTGCAAGAACTGAGTGCAG 6572  
 1085 upheAsnArgTyrGlnGlyThrLeuSerValTyrGlnSerAsnIleSer 1102  
 6573 GAAATAGCTGCTAGCTGAGAGAGTTCAGTACTGTCGAAATTAAGA 6622  
 1102 yshLysGlnGluLeuLysLeuSerGlnTyrLysAspIleGluLysArg 1118  
 6623 GCAGCAAAAGATCAATTGAAG..... 6642  
 1119 TyrThrAsnGlnPheLeuGlnLeuLysThrThrGluMetAlaAsnLysAs 1135  
 6643 .....GAGCTCACACTAGAAATAGTGA..... 6666  
 1135 pleuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPheHis 1152  
 6667 .....TTGAAGAGAGCTTAGATTCATGCATGCACA 6694  
 1152 erMetLysMetGluGluIleAsnLysIleIleLysGluLeuTyrGlnGln 1168  
 6695 AAGACCAAGTGGAAAGAGAGAGAGTGAAGAGAAATAGCTGAATAT 6744  
 1169 ThrTyrArgGlyGlnAspIleAsp.....TyrIleSerIleAsnSerAs 1183  
 6745 CAGTACAGGCTTCATGACCTGAAGAAACACAGCTTTCCTTTGGA 6794  
 1183 pSerGluGlyAlaGlyThrArgSerTyrArgValIleMetGlnT 1200  
 6795 CACA.....AACACACATGATGAAGAAATCCAGA 6826  
 1200 hGlyAspAlaGluLeuGluMetArgGlyArgCysSerAlaGlyGlnLys 1216  
 6827 CATACCGAGAGAAATGACTCTAAAGAGAAATGCTCAGTTCCAGAGAG 6876  
 1217 ValLeuAlaSerLeuIle 1222  
 6877 CTGGAGATAGACCTTTTA 6894

seq\_name: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:US-08-353-700-2

seg\_documentation\_block:  
 ; Sequence 2, Application US/08353700  
 ; Patent No. 5599919  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YEN, TIMOTHY J.

APPLICANT: RATTNER, JEROME B.  
 TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
 TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
 TITLE OF INVENTION: AND METHODS OF USE  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSES: DANN, DOREMAN, HERRELL AND SKILLMAN  
 STREET: 1601 MARKET STREET, SUITE 720  
 CITY: PHILADELPHIA  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103-2307  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/353,700  
 FILING DATE: 09-DEC-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: REED, JANET E.  
 REGISTRATION NUMBER: 36,252  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 563-4100  
 TELEFAX: (215) 563-4044  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10136 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: HUMAN  
 US-08-353-700-2

alignment\_scores:  
 Quality: 439.50 Length: 1256  
 Ratio: 0.668 Gaps: 49  
 Percent Similarity: 52.389 Percent Identity: 19.904

alignment\_block:  
 US-09-538-396-2 x US-08-353-700-2 ..

Align seg 1/1 to: US-08-353-700-2 from: 1 to: 10136

51 CysThrGlyGluLeuProAsnSerArgSerGlyHisThrPheValH 67  
 5619 TGCTCTGGGAGACAGTCCACGATACCAT.....TA 5650  
 67 aaSPProLysValAlaGlyGluThrGluThrLysGlyGlnIleLysLeu 84  
 5651 TGAGCT.....CCAGGGAGAGATTAACCCAGGCTCTTCAGAA... 5690  
 64 rghelyshThrAlaAlaGlyLysAspValValCysIleArg.....Ser 98  
 5691 .....TGCAATTCGTGAATGTCA 5708  
 99 PheGlnLeuThrGlnLysAlaSerLysMetGluPhe.....Lys 111  
 5709 TTTCCTGCTCCATATGCTTGTACTATGATTCCTCTGGGAATCAGCA 5758  
 111 salalIleGluSerValLeuGlnThrIleAsnProHisThrGlyGlyGly 128  
 5739 AGATATCCATTAATCTTCACTCGGGGTAAAGACATCAATAGAGAAAT 5808  
 128 aLcysLeuSerTyrArgCysAlaAspMetAspArgGluLeuProAlaLeu 144



```

707 GluAspGluPheValLysGlnArgMetGlnAsnSerThrAlaGln 723
    |||:|||||
738 GAGGAAGAG.....CATCAGCTGGAATAGCATTAAGAAAGCT 7375
723 uArgSerLysAlaLeuAlaMetGluSerSerAsnAlaGlnAlaLeuPheG 740
    |||:|||||
7376 GAGAGCCCGCTGAGAACCTGATGAAGAACACAGCTC...TGCTCTTAC 7422
740 LngLysAspLysLeuArg.....ThrIleTyAspAlaTyrVal 753
    |||:|||||
7423 AACAACTGAAGGAAGAGTGCATCATCATCAGATTACTTAAGGATGAGTGTG 7472
754 LysLeuValGlnGluThrIleProLeuAlaGlnLysAsnLeuSnglnH 770
    |||:|||||
7473 GAGAACCTTGAAGAGAGCTAGAGATAGCCAGACAAAC...CAAGAGCA 7519
770 sLeuAlaAspGluSerGlnLysAlaGlnAlaPheAspAspLeuGlnLys 787
    |||:|||||
7520 TGCAGCTCTTGAGGCGAGCAATTCGCAAA..... 7547
787 alLeuAlaHisValGlnMetAspArgAspAlaValGlnAlaLeuGln 803
    |||:|||||
7548 ....GAGAGCTAGAGACCCCTAAAGCAAAATAGAGGATGACCCAA 7592
804 ProThrAspThrIleAspArgHisVal..... 812
7593 AGTGTAGAGAGCTCTGGAATAGATGTGTACTATAAGTCAGCAAAAGA 7642
813 .....HisGluIleGlnGlnLeuValLysGlnValGlnAspLeuG 826
7643 AAATCTGACAAATGAATTCACAAAGCAAGCGCAATATCTGAATTAG 7692
826 LysAlaLeuAspSerSerGlyArgGlyVal.....LysSerLeu 839
7693 AA...ATAATAATTCATCATTTTGAATAATTTTGCAGAAAGAGCAAA 7739
840 GlnGluIleGlnLeuGlnLeuAsnPheLeuGlnArgThrArgAspThrIle 856
7740 GAGAAAGTACAGATG.....AAGAAATCAACAC 7771
856 uLeuValGlnValAspAspLeuArgAspGlnHisArgMetLeuAsnGln 873
7772 TGCATGAGGATG.....CTTCAACACATTAAGAGCTCAATGAGA 7815
873 sPheSerSerAlaGlnValArgTrpHisAsnAlaArgGlnGlu...Lys 888
7816 GAGTGCACCCCTG.....CATATGACCAAGAGAGCTGTAAAG 7853
889 ValLysAlaSerSerIleLeuGlnArgPheGlnLysSerGlnGlnLys 905
7854 GCCAAAGACAGAAAT..... 7868
905 uValLeuLeuAlaGlnGlnLysGlnGlnLeuValGlnLysLysLeu 922
7869 .....CTTAGTAGTCAAGTAGAGTGTCTTGAATTGAAGAGCTCAG 7911
922 euGlnGlnSerLeuAspPro.....LeuSerLys 931
7912 TGCCTCAAGGCTTGATGAGGCCAAATAATATATGTTTGTTCATCT 7961
932 GlnLysGlnSerLeuLeuGlnGlnLysArgAlaLeuLysGlnLysLeu 948
7962 TCACGAGAAAGGCTCATTTCAAGAGTAGAGATGGCAGACAGAACTGGA 8011
948 pGlnGluTyrHisGlnLeuAlaGlnArgLysArgGlnPheGlnGlnLys 965
8012 GAAGAGAGGTGAAGAAATCAGTAGACTGAATAATCAATTCAGAGCAAG 8061
965 euAspAlaLeuGlnLysArgLeuAsnMetLysIleLysGlnTyrLeuAspSer 981
8062 AGCAGCTTGTTC..... 8072
982 LysLysAsnGlnLysLeuLysGlnLeuGlnGlnLysArgHisValLeuGlyshi 998

```

```

8073 .....CTTAACCTGCCAGTGGAGAGAGAGACCAACTTTGGA 8113
998 sSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgIleSerAlaGln 1015
8114 GGACCAA.....AACTTAGAAGTGAAGAAATGACGGTGGAT 8151
1015 euAsnLysSerLysGlnLeuLeuGlnGlnGlnGlnLysArgAsn 1031
8152 TGGAGCAGAAATCCAGTCTCAATCCAA.....AATGCCCT 8192
1032 IleAspAsnLeuLys.....TyrArgLysThrLysAl 1043
8193 TTGCAGACACATTTAGAAAGTCTCAGAGTCTTACAGAAATCTAGAGAA 8242
1043 AspValGlnGlnLeuThrArgAspIleGlnSerLeuGlnGlnArgLeu 1060
8243 TGACCTTGAATTTGACAAATAATGACAAATAGTCTTTGTGAAAAA... 8288
1060 euSerIleGlnSerLeuSerAlaIleGlnAlaAspLeuLysArgHisSer 1076
8289 .....GTAAACAAATGACTGCAGAAAGAACTGAGCTGACAGAGGAATG 8333
1077 GlnGlu.....LysGlnArgLeuAsnSerG 1085
8334 CATAGATGGCAGCAAGAAACAGCAGACTGCAGAGAACTGAGTGGAGA 8383
1085 uPheAsnArgTyrPheGlnLysThrLeuSerValTyrGlnSerAsnIleSer 1102
8384 GAAATAATAGCTAGCTGAGAGTGTGCAAGTTACGTTGGAAGAAATAAGA 8433
1102 yHisLysGlnGlnLeuLysLeuSerGlnTyrLysAspIleGlnLysArg 1118
8434 GCAGCAAAAGATCAATTGAG..... 8453
1119 TyrThrAsnGlnPheLeuGlnLeuLysThrThrGlnMetAlaAsnLysAs 1135
8454 .....GAGCTCACACTAGAAATATGGA..... 8477
1135 pLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPheHis 1152
8478 .....TTGAAGAAAGACCTAGATGATGATGCAC 8505
1152 eMetLysMetGlnGlnLysAlaLysIleIleLysGlnLeuTrpGlnGln 1168
8506 AAGACCAGGTGGAAAGAGAGAGAGTGAAGAGAAATAGCTGAATAT 8555
1169 ThrTyrArgGlnGlnAspLeu.....TyrIleSerIleAsnSer 1183
8556 CAGCTACGGCTTCATGAGCTGAAAGAAACACAGAGCTTGTGTTTGA 8605
1183 pSerGlnGlnAlaGlnLysArgSerTyrTyrArgValIleMetGln 1200
8606 CAA.....AACAAACAGTATGAGTAGAGAAATCAG 8637
1200 hGlnLysAlaGlnLeuGlnLeuMetArgLysArgCysSerAlaGlnLys 1216
8638 CATACCGAGAGAAATGACTTCTTAAGAAAGATGTCTCAGTTCACAGAG 8687
1217 ValLeuAlaSerLeuIle 1222
8688 CTGAGATAGACCTTTTA 8705

```

seq\_name: /cgn2\_6/ptodata/2/ina/PCrus\_COMB.seq:PCr-US95-16216-2.

seq\_documentation\_block:

Sequence 2, Application PC/TUS9516216

GENERAL INFORMATION:

APPLICANT: Yen, Timothy J.

APPLICANT: Ratner, Jerome B.

TITLE OF INVENTION: Nucleic Acid Encoding a Transiently

NUMBER OF SEQUENCES: 4 Expressed Kinetochores Protein, and Methods of Use